

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 05:26:48 ; Search time 2436.75 Seconds  
(without alignments)  
16488.745 Million cell updates/sec

Title: US-09-786-474-1

Perfect score: 1920

Sequence: 1 atgctgcgcctcttgcac.....gctcctaccggctgagcctt 1920

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

rched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_scs:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pi:\*

26: em\_ro:\*

27: em\_scs:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description |
|------------|-------------|-------|--------|----|-------------|
|------------|-------------|-------|--------|----|-------------|

ALIGNMENTS

RESULT 1

AB029550

LOCUS

DEFINITION

AB029550

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AB029550 3825 bp DNA linear BCT 23-MAY-2000  
Corynebacterium glutamicum genes for ltsA, ORF1, complete cds.

AB029550.1 GI:6714539

ORF1; ltsA.

Corynebacterium glutamicum (strain:KY9611) DNA.

Corynebacterium glutamicum

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Corynebacteriaceae;

Corynebacterium.

1 (sites)

Hirasawa, T., Wachi, M. and Nagai, K.

A mutation in the corynebacterium glutamicum ltsA gene causes

susceptibility to lysozyme, temperature-sensitive growth, and

L-glutamate production

J. Bacteriol. 182 (10), 2696-2701 (2000)

20245524

2 (bases 1 to 3825)

Wachi, M. and Hirasawa, T.

Direct Submission

Submitted (03-JUL-1999) Masaaki Wachi, Tokyo Institute of

Technology, Department of Bioengineering; 4259 Nagatsuta Midori-ku,  
Yokohama, Kanagawa 226-8501, Japan (E-mail: mwachi@bio.titech.ac.jp,  
Tel: 81-45-924-5770, Fax: 81-45-924-5820)

#### FEATURES

Location/Qualifiers  
1..3825  
/organism="Corynebacterium glutamicum"  
/strain="KY9611"  
/db\_xref="taxon:1718"  
815..2737  
/gene="ltsA"  
815..2737  
/gene="ltsA"  
/codon\_start=1  
/transl\_table=11  
/product="ltsA"  
/protein\_id="BAA89484.1"  
/db\_xref="GI:6714540"  
/translation="MGLLGILITANGNAEAFVPALERALPCMRHRGDDAGTWHADADA  
AFGNRUSIIDIAHSHQPLRWGPADEPRDYAMTNGEYTYNVELRKELSDLGTYFNFS  
GDGPEIIVGFHHGSEVVEHLRGMFGIAINDTKESLFLARDQFQIKRPLEYATTEHGT  
VFSSEKKTILMEAMENMLDLGLDKRTIEHYVDLQVPEPDTLHAQISRLSGCTATVR  
PGGLEOKRYFKPQPKVVKVKGEOQLDFRIAQVLEDSVEKHMRAQVTVGSFSLGGI  
DSTAIAPLAKRHPDLDTFTGFEREGYSEVDVAESAALGAHIIKVIISPEYANA  
IPKIMWLDPPADPSLVPLIFVAAEARKHKVVKVLSGGADELFGGTTIYKEPUSLAP  
FEKIPSLRKLGLSKVLDPDMKGKSLLERGSMTEERYVGNARSFNFMQRMVPM  
AKRMDHREVTAPIYAQSRNPDPVARHQLDLFTWMRGDILVKADKINMANSLRLVP  
FLDEKVFVAETIPYDLKIANGTTKYALRRALQIIVPHVLRKKLGFPPVPMRHLWLAG  
DELFGWAQDITKESGTEDIENKQAVLMDLNEHRDGVSDHSRRLTWLVSFMYWHGIFVE  
NRIDPQIEDRSYPVEL"  
complement(2882..3226)  
/gene="orf1"  
complement(2882..3226)  
/gene="orf1"  
/codon\_start=1  
/transl\_table=11  
/product="orf1"  
/protein\_id="BAA89485.1"  
/db\_xref="GI:6714541"  
/translation="MTAPSTNTGVIILTESAASKAKALIDQGRDDLRLIRAVQPGGCS  
GLRYLVFDORTLDGDKREDIVGVRLVVDKMTPEYLLGAQIDFADTIEQQGFTIDNPN  
AGSSCAGDSFN"  
BASE COUNT 881 a 1031 c 993 g 920 t  
ORIGIN  
Query Match 100.0%; Score 1920; DB 1; Length 3825;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atgtgcgccctcttctggcatattgactcaaatgggaacgctgaagcattcgttctctgca 60  
815 ATGTGCGGCCTTCTTGGCATATTGACTCAAAATGGGAACGCTGAAGCATTCGTTCTCTGCA 874  
QY 61 ctgagcgggctctgcatgcatgcgcacacgctggtctctgacgatgcggcacttggcat 120  
Db 875 CTCGAGCGGGCTTCCCATGATGCGCCACCGCTGCTCTGACGATGCCGGCACTTGGCAT 934  
QY 121 gacgcccgtgacgcttggattcaaccqctctccatcattgatattgcacactccacac 180  
Db 935 GACGCCGATGCAGCGTTTGGATTCAACCGCTCTCCATCATTTGATATTCACACTCCAC 994  
QY 181 caaccactgcttggggacctgcggatgaaccgacgctcagcaatgacttcaacggt 240  
Db 995 CAACCACTGCGTTGGGGACCTGCGGATGAACCGCACCGCTACGCAATGACTTCAACGGT 1054  
QY 241 gagactcaactcagttgagctgcgttaagagctctctcgatttgggatatacccttaat 300  
Db 1055 GAGATCTACACTAGCTTGACGTGCGTAAAGAGCTCTCGGATTTGGGATATACCTTTTAA 1114  
QY 301 acttctgcgatggcgagcccaattgttgcggtttccaccactggggcgagtcctggtgc 360  
Db 1115 ACTTCTGCGGATGGCGAGCAATATTGTTGTCGTTTCCACCACCTGGGGCGAGTCCGTGTC 1174

QY 361 gagcatctccgcggaattgttccggcattgccatttgggatacaaaagaaagtcgcttttc 420  
Db 1175 GAGCATCTCCGCGGAATGTTCCGCATTGCCATTGGGATACAAAGAAAGTCGCTTTTC 1234  
QY 421 ctgcgcgctgacgttctggcatcaagccactgttctacgaaccacccagagatggcacc 480  
Db 1235 CPTGCGCGTGATCAGTTCGGCATCAAGCCACTGTTCTACGCCAACCCAGCAGATGGCACC 1294  
QY 481 gtgttctctcagaagaagaaccatcttggagatgqcccgagagagatgaatctagatctg 540  
Db 1295 GTGTTCTCTCAGAGAGAGAGACCATCTTGGAGATGGCGGAGAGATGATCTAGATCTG 1354  
QY 541 ggccttgaagcgcaccattgagcactacgttggaccctgcagctacgtgcccagagccagat 600  
Db 1355 GGCCTTGTATAGCGCACCATTTGAGCACTAGCTGAGCTGACGTACGTGCTGCCGAGCAGAT 1414  
QY 601 acccttcacgcgcagatttcccgccttggagtcagctgcacccgcaacagttcgttcgggac 560  
Db 1415 ACCCTTTCACGCGCAGATTTCGCGCTTGGAGTCAGCTGCACGCCAACAGTTCGTCCGGGC 1474  
QY 661 ggcagcctggaacagaagcgttacttcaagcctcagttcccaagtacagaagtcgtaaaag 720  
Db 1475 GCGAAGCTGGAAACAGAGCGTTACTTCAAGCCTCAGTTCCCACTACAGANAGTCTGAAG 1534  
QY 721 ggttaaggagcaggacctcttcogctgcattgcccagggtgttggagagatagctcgaagaag 780  
Db 1535 GGTAAAGGAGCAGGACCTCTTCGATCGCATTGCAGAGTGTTCGAGGATAGCGTTCGAAAA 1594  
QY 781 catatgctgcgcagctgacccgtaggctcgttcccttccggcgcgagctgactcaaccgcga 840  
Db 1595 CATATGCGTGGCGAGCTGACCGTAGGCTCGTTCCTTCGCGCGGCAATTGACTCAACCGCA 1654  
QY 841 attgcgcgccttggaaaagcccaaacctgcacctgccttccaccacccggttttcgag 900  
Db 1655 ATTGCGCGCGCTTGAAGAGCGCCACAACTGACCTGCTCACCCTTCAACCCCGTTTCGAG 1714  
QY 901 cgtgaagcctactcgcggaggtcgatgtgctgcggagtcgcgcgctgcgattggcgctgag 960  
Db 1715 CGTGAAGGCTACTCGGAGGTGCGATGTGGCTGGGAGTCCGCGCTGCGATTGGCGCTGAG 1774  
QY 961 cacatcgtgaagattgtctcgctgaggaaatacgcccaacgcgagattcctaagatcatgtg 1020  
Db 1775 CACATCGTGAAGATTGTCTCGCCTGAGGAATACGCCAACGCGATTCCTTAAGATCATGTGG 1834  
QY 1021 tacttggatgacctgtgactgaccccatcattgttccgcgctgacttctgtgscagcgaaa 1080  
Db 1835 TACTTGGATGATCTGTAGCTGACCCATCATTTGGTCCGCTGTACTTCTGCGCAGCGNA 1894  
QY 1081 gcaogtaagcagctcaaggttctgtctgtgagggcgagagtgagctgttcggtgga 1140  
Db 1895 GCACGTAAGCAGCTCAAGGTTGTGCTGTGGCGAGGGCGCAGATGAGCTGTTCGTTGGA 1954  
QY 1141 tacaccatttcaaaagacccgctatcgttctccatttgagaagatcccttcccacta 1200  
Db 1955 TACACCAATTTACAAAGACCGCGTATCGCTTGTCTCCATTGAGAAATCCCTTCCCCACTA 2014  
QY 1201 cgtaaagccttgggaagctcagcaaggttcttcagacgcgcgatgaagggcaagtccctt 1260  
Db 2015 CGTAAAGCCCTGGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAGGGCAAGTCCCTT 2074  
QY 1261 cttgagcgtgctccatgacattggaagagcgctactacgcgaacgctcgtctcctcaat 1320  
Db 2075 CTTGAGCGTGGCTCCATGACCATGGAGAGGCGTACTACGGCAACGCTCGCTCTTCAAT 2134  
QY 1321 ttcagcagatgcaacgcgttatttccatgggcaagcgcaatgggacacccgcggaagtc 1380  
Db 2135 TTGAGCAGATGCAACGCGTTATTTCCATGGGCAAGCGGAAATGGGACACCCGCCGAAGTC 2194  
QY 1381 actgcaccgatctacgcacaaatcccgcgaactttgatccagtagccccgcgatgcaacactg 1440  
Db 2195 ACTGCACCGGATCTACGCACAATTCGGCAACTTTGATCCAGTAGCCCGCATGCAACACCTG 2254  
QY 1441 gatctgttcaactggatgcgcggcgacaactcctggttcaaggctgacaagatcaaacatggcg 1500

|  |   |  |  |         |                            |
|--|---|--|--|---------|----------------------------|
| Db   | 2255  |  | GATCTGTTACCTTGGATGCGCGGACATCCTGGTCAAGGCTGACAAGATCAACATGGCG   | 2314    |                            |
| Qy   | 1501  |  | aactcccttgagctgaggttccattcttgataaagaaagttttcaaggttgagagacc   | 1560    |                            |
| Db   | 2315  |  | AATCCCTTGAGCTGGAGTTCCATTCTTGGATAAAGGAAGTTTCAAGGTTGCAGAGACC   | 2374    |                            |
| Qy   | 1561  |  | attccttaacgatcgaagtgcacaggttaccaacaagtacgcgtgcagggcactc      | 1620    |                            |
| Db   | 2375  |  | ATTCCCTTACGATCTGAAGATTGCCACGGTACCACCAATGACCGCTGCGCAGGGCACTC  | 2434    |                            |
| Qy   | 1621  |  | gagcagatttccgcctcaagttttgacccgcaagaagctgggcttccctgttcccatg   | 1680    |                            |
| Db   | 2435  |  | GAGCAGATTGTTCCGGCTCACGTTTTCACCGCAAGAAGCTGGGCTTCCCTGTTCCCATG  | 2494    |                            |
| Qy   | 1681  |  | cgcactggcttcgagcagatgagctgttcggttggtggcgagacaccattaaagaaatcc | 1740    |                            |
| Db   | 2495  |  | CGCACGTGGCTTCGGCGGATGAGCTGCTCGTGGCGCAGGACACCAATTAAGGAATCC    | 2554    |                            |
| Qy   | 1741  |  | ggtactgaagatatcttcaacaagcaggtgtgctgatatgtgaacgagcacccgcgat   | 1800    |                            |
| Db   | 2555  |  | GGTACTGAAGATATCTTCAACAAGCAGGCTGTGCTGGATATGCTGAACGACACCGGAT   | 2614    |                            |
| Qy   | 1801  |  | ggcgtgcagatattccctgcagctgagctgttcgtcatattatggtgtggcacggc     | 1860    |                            |
| Db   | 2615  |  | GGCGTGTACATCATTTCCCTGCAGCTGTGACCTGTCTGTCATTTATGTTGTGCACGGC   | 2674    |                            |
| Qy   | 1861  |  | atttttggaaaaccgcatgtatccacagatgagacgcctcctaccggctcgagctt     | 1920    |                            |
| Db   | 2675  |  | ATTTTGTGGAAAACCGCATTTGATCCACAGATTGAGGACCGCTCCTACCCGGTCGAGCTT | 2734    |                            |
| RESULT 2   |   |  |  |         |                            |
| AX122490   |   |  |  |         |                            |
| LOCUS  | AX122490  | Sequence 2406 from Patent EPI108790.                         |  | 1920 bp | DNA linear PAT 11-MAY-2001 |
| DEFINITION   | AX122490  |  |  |         |                            |
| ACCESSION  | AX122490  |  |  |         |                            |
| VERSION  | AX122490.1  | GI:14039816  |  |         |                            |
| KEYWORDS   |   |  |  |         |                            |
| SOURCE   |   |  |  |         |                            |
| ORGANISM   | Corynebacterium glutamicum.   |  |  |         |                            |
|  | Corynebacterium glutamicum.   |  |  |         |                            |
|  | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;   |  |  |         |                            |
|  | Actinomycetales; Corynebacterineae; Corynebacteriaceae;   |  |  |         |                            |
|  | Corynebacterium.  |  |  |         |                            |
| REFERENCE  | 1 (bases 1 to 1920)   |  |  |         |                            |
| AUTHORS  | Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A. |  |  |         |                            |
| TITLE  | Novel polynucleotides   |  |  |         |                            |
| JOURNAL  | Patent: EP 1108790-A 2406 20-JUN-2001;  |  |  |         |                            |
|  | KYOWA HAKKO KOGYO CO., LTD. (JP)  |  |  |         |                            |
| FEATURES   | Location/Qualifiers   |  |  |         |                            |
| source   | 1..1920   |  |  |         |                            |
|  | .organism="Corynebacterium glutamicum"  |  |  |         |                            |
|  | /db_xref="taxon:1718"   |  |  |         |                            |
| BASE COUNT   | 422 a   | 534 c  | 528 g  | 436 t   |                            |
| ORIGIN   |   |  |  |         |                            |
| Query Match 99.2%; Score 1904; DB 6; Length 1920;              |   |  |  |         |                            |
| Best Local Similarity 99.5%; Pred. No. 0;                      |   |  |  |         |                            |
| Matches 1910; Conservative 0; Mismatches 10; Indels 0; Gaps 0; |   |  |  |         |                            |
| Qy   | 1   | atctgcggcctcttgccatattgactgcaaatgggaacgtgaagcattcgttctcgca   | 60   |         |                            |
| Db   | 1   | ATGTGCGGCCCTCTTGGCATATTGACTGCAATGGGAACGCTGAAGCATTTCTGTTCTGCA | 60   |         |                            |
| Qy   | 61  | ctcagcggccttgccatgcacgcaccgttggtctctgacgatgcgcgaccttgcat     | 120  |         |                            |
| Db   | 61  | CTGAGCGGGCCCTTGCCATGCATGCGCCACCGTGGTCTTGACGATGCGGGCATTGGCAT  | 120  |         |                            |
| Qy   | 121   | gagccgatcagcgtttggattcaaccgcctctccatcttgcatttgcacactccac     | 180  |         |                            |
| Db   | 121   | GAGCCGATCGAGGTTGGATTCAACCGCTCTCCATCATTTGATATTGACACTCCAC      | 180  |         |                            |

|    |      |  |      |  |
|----|------|--|------|--|
| Qy | 181  | caaccactgcttgggaacctgcgagtaaacccgacccgtctacgaatgactttcaacggt     | 240  |  |
| Db | 181  | CAACCACCTGCTTGGGACCTGCGGATGAACCCGACCGCTACGCAATGACTTTCAACGGT      | 240  |  |
| Qy | 241  | gagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatatacctttaa    | 300  |  |
| Db | 241  | GAGATCTACAACCTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTGGGATATGCTTTAAT     | 300  |  |
| Qy | 301  | actctggcgataggcgagccaaattgttgcggtttccaccactggggcgagtcgtagtgc     | 360  |  |
| Db | 301  | ACTTCTGGCGATGGCGAGCCAAATGTTGTCGGTTTCACCACCTGGGCGGAGTCGGTGGTC     | 360  |  |
| Qy | 361  | gagcatctccgcggaattgttcggcatttgggatacaaaagaaagagtcgcttttc         | 420  |  |
| Db | 361  | GAGCATCTCCGCGGAATGTTTCGGCATTTGCCATTTGGGATACAAAGGAAAGTCGCTTTTC    | 420  |  |
| Qy | 421  | cttgcgctgatacagttccggcataaagccaactgtctctacgcaaccaccgagcatggcaac  | 480  |  |
| Db | 421  | CTTGC CGCTGATCAGTTTCGGCATTTAAGCCACTGTGTTCTACGCAACACCGAGCATGGCACC | 480  |  |
| Qy | 481  | gtttctctcagagaagaagaccatcttgagatggccgagagagatgaatctagatctg       | 540  |  |
| Db | 481  | GTGTTCTCTCAGAGAAAGAGACCATCTTGAGATGGCCGAGGAGATGAATCTAGATCTG       | 540  |  |
| Qy | 541  | ggccttgataagcgcaccattgagcactacgttggaacctgcagtacgtgcccgcagccagat  | 600  |  |
| Db | 541  | GGCCTTGATTAAGCGCACCATTTGAGCACTAGTGGACTTGCAGTACGTGCCGAGCCAGAT     | 600  |  |
| Qy | 601  | accttcaecgcgagatttcccgccttggagtcagggtgcaccgcgcaacagttctgcgggc    | 660  |  |
| Db | 601  | ACCTTTCACGCGCAGATTTCGCCGCTTGGAGTCAAGGCTGACCCGCAACAGTTTCGTCGGGC   | 660  |  |
| Qy | 661  | ggcaagctggaacagaagcgttacttcaagcctcagttcccagtcacagaagagtcgtaaa    | 720  |  |
| Db | 661  | GGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTCAGTTCCCAAGTACAGAAGTCTGAAG      | 720  |  |
| Qy | 721  | ggtaaaggagcaggacctcttcgcagtcgattgccagagtggtggaggtagcgtcgcaaa     | 780  |  |
| Db | 721  | GGTAAGGAGCAGGACCTCTTCGATCGCATTTGCCAGGTGTTGGAGGATAGCGTCGAAAG      | 780  |  |
| Qy | 781  | catatgctgcgacgtgacgtaggtgctgtctcttccgcggcgatctgactcaaccgca       | 840  |  |
| Db | 781  | CATATGCTGCCGACGTTGACGTTAGGCTCGTTCTTTCCGGCGGCAATGACTCAACGCA       | 840  |  |
| Qy | 841  | attgcgcgcttcaaaagcgcacacccctgacctgacctcaccctcaccacgcgggttcg      | 900  |  |
| Db | 841  | ATTGCGGCGCTTGCAAGCGCCACACCCCTGACCTGCTCACCTTCACCCACCGGTTTCGAG     | 900  |  |
| Qy | 901  | cgtgaaggttactcggaggttcgattggtgctgcggagttccgcgcgtgcgattggcgctg    | 960  |  |
| Db | 901  | CGTGAAGGCTACTCGGAGGTGCGATGTCGCTGCGGAGTCCGCCGCTGCGATTGGCGCTGAG    | 960  |  |
| Qy | 961  | cacatgctgaagattgtctgcgctgaggaatacgcgaacgcgattcctaagatcatgtgg     | 1020 |  |
| Db | 961  | CACATCGTGAAGATTGTCGCTGAGGAATACGCCAACGCGATTCTCTAAGATCATGTGG       | 1020 |  |
| Qy | 1021 | tacttgatgatcctgtagctgaccteatattggtcccgctgtactctcgtaggcaggaa      | 1080 |  |
| Db | 1021 | TACTTGGATGATCCTGTAGCTGACCCCATCATTTGGTCCCGCTGTACTTTCGTGGCAGCGAA   | 1080 |  |
| Qy | 1081 | gcagtaagcacgtcaaggttgctgtctcggagagggcgagatgagctgttcggtgga        | 1140 |  |
| Db | 1081 | GCAGTAAGCACGTCGAAGGTTGCTGTCTGCGGAGGGCGCAGATGAGCTGTTCGGTGGGA      | 1140 |  |
| Qy | 1141 | tacaccattacaagagccgctatcgcttgcctcatttgagaagatcccttccccacta       | 1200 |  |
| Db | 1141 | TACACCATTTACAAGGAGCCGCTATTCGCTTGCATTTTGAGAAGATCCCTTCCCCACTA      | 1200 |  |
| Qy | 1201 | cgtaaagcctcggaaagctcagcaaggttctgccagacggcattgaagggcaagtcacct     | 1260 |  |
| Db | 1201 | CGTAAAGCCTGGGAAAGCTCAGCAAGGTTTCTGCCAGAGCGCATGAAGGGCAAGTCCCTT     | 1260 |  |

Qy 1261 cttgagctggctccatgaccatggaagagcgctactacggcaacgctgcgtcccttcaat 1320  
|||||  
Db 1261 CTTGAGCTGGCTCCATGACCATGGAAGAGCGCTACTAGGGCAAGCTCGCTCCTTCAT 1320  
  
Qy 1321 ttcgagcagatgcaacgcgtttatccatggtggaagcggaatggaacacgcggaagtc 1380  
|||||  
Db 1321 TTCGAGCAGATGCAACGCGCTTATTCATGGGCAAGCGCAATGGACCCGCGAAGTC 1380  
  
Qy 1381 actgacagcatcagcacaatcccgcgaactttgcatccagtagcgcgcgatgcaaacactg 1440  
|||||  
Db 1381 ACTGCGCCGATCTAGCAGAGTCCCGCAACTTTGATTCAGTAGCCGATGCAACACCTG 1440  
  
Qy 1441 gatctgtcacctggatgcgcggcgacatccctggtcaaggtgacaagatcaacatggcg 1500  
|||||  
Db 1441 GATCTGTTACCTGGATGGCGGCGACATCTCTGGTCAAGGCTGACAAGATCAACATGGCG 1500  
  
Qy 1501 aactcccttgagctgcgagttccattctctggaagaagttttcaagttgagagacc 1560  
|||||  
Db 1501 AACTCCCTTGAGCTGCGAGTTCCATTTCTGGATAAGGAAGTTTTCAGGTTGCAGAGACC 1560  
  
Qy 1561 attccttcagatcgaagattgcgaacgtagtaccaccaagtagcgcgtcgcgagggcactc 1620  
|||||  
Db 1561 ATTCCTTAGCACTGAAGATTGCCAACGGTAGCCACCAAGTAGCCGCTGGCAGGGCACTC 1620  
  
Qy 1621 gagcagattgttccgcctcactgcttttgacccgcaagaagctgggcttccctgttccccatg 1680  
|||||  
Db 1621 GAGCAGATTGTTCCGCCCTCACGTTTTTGACCGCAAGAAGCTGGGCTTCCC GTTCCCATG 1680  
  
Qy 1681 cgcactggcttgcggcgatgagctgttcggttggcgcgaggacacattaaagaaatcc 1740  
|||||  
Db 1681 CGCCACTGGCTTGGCGGCGATGAGCTGTTCGGTGTGGCGCAGGACACCATCAAGGAATCC 1740  
  
Qy 1741 ggtactgaagatatctcaacaagcagctgtgctgggatgctgaacgagcaccgcgat 1800  
|||||  
Db 1741 GGTACTGAAGATATCTTCAACNAGCAGGCTGTGCTGGATATGCTGAACGAGCACCCECAT 1800  
  
Qy 1801 ggcgtgcagatcattccgctgactgtggactgtctgtcatattatggtggcgacggc 1860  
|||||  
Db 1801 GCGGTGCAGATCATTCGCTGCAGCTGGAGCTGTCTGTGCTATTTATGTTGGCGACGCGC 1860  
  
Qy 1861 atttttggaaaacccgatgatccacagattgagaccgctctcaccggctgagatt 1920  
|||||  
Db 1861 ATTTTGTGGAAAACCGCATGTATCCAGATTTGAGGACCGCTCTTACCCAGTCGAGCTT 1920  
  
RESULT 3  
AX127150 349980 bp DNA linear PAT 11-MAY-2001  
LOCUS Sequence 7066 from Patent EP1108790.  
DEFINITION AX127150 AX114121  
ACCESSION AX127150.1 GI:14041138  
SOURCE  
Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;  
Corynebacterium.  
1 (bases 1 to 349980)  
Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,  
Yokoi H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.  
Novel polynucleotides  
Patent: EP 1108790-A 7066 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
Location/Qualifiers  
1. 349980  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3.309.400) split in 11, seq 7066  
2.100.001 2.449.980"  
BASE COUNT 80724 a 98367 c 90490 g 80399 t  
ORIGIN

Query Match 99.2%; Score 1904; DB 6; Length 349980;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1910; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
Qy 1 atgtgcggccttcttggcatattgactgcaaatgggaacgctgaagcattocttctgca 60  
|||||  
Db 228516 ATGTGCGGCTTCTTGGCATATTGACTCAAAATGGGAACGCTGAAGCATTCGTTCTCTGCA 228575  
  
Qy 61 ctcgagcgggcttgccatgcatgcccacgctggtctctgacgatgcccgaacttggcat 120  
|||||  
Db 228576 CTCGAGCGGGCTTCCCATGTCATGCGCACCGCTGCTCTGACGATGCGCGCACTTTGGCAT 228635  
  
Qy 121 gacgcgcgatgacgctttggattcaacgcgcctctccatcattgatattgacacatcccccac 180  
|||||  
Db 228636 GACGCGGATGACGCGTTTGGATTCAACCGCTCTCCATCATTTGATATTGCACACTCCAC 228695  
  
Qy 181 caaccactgcgttggggaccctgcggatgaacccgaccctacgcgaatgactttcaacggt 240  
|||||  
Db 228696 CAACCACCTGCGTTGGGGACCTGCGGATGAACCGCACCGCTACGCAATGACTTTCAACGCT 228755  
  
Qy 241 gagatctcaactacgtctgagctgcgttaagagctctcggatttgggatataccttbaat 300  
|||||  
Db 228756 GAGATCTCAACTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTGGGATATGCCCTTTAAT 228815  
  
Qy 301 acttctggcgatggcgagcccaattgttcggtttccaccactggggcgagtcctgtggtc 360  
|||||  
Db 228816 ACTTCTGGGATGGCGAGCCCAATTGTTGCTGTTTCCACCCTGSGGCGAGTCCGTGGTC 228875  
  
Qy 361 gagcatctccgcgggaattctcggcaattgccatttgggatacaaaagaaagtcgcttttc 420  
|||||  
Db 228876 GAGCATCTCCGCGGAATGTTCCGGCATTTGCCATTTTGGGATACAAAGGAAAGTCGCTTTTC 228935  
  
Qy 421 cttcgcgtgatcagattcgcgcgatcaagcactgttctacgcaacccaccagcagcatggcacc 480  
|||||  
Db 228936 CTTGCGGTGATCAGTCTCGGCATTAAGCCACTGTCTTACGCAACCCAGCATGGCAC 228995  
  
Qy 481 gttcttctctcagagaagaagaccatcttggagatggcgaggagatgaattagatctg 540  
|||||  
Db 228996 GTGTTCTCTCAGAGAAGAAGACCATCTTTGGAGATGGCGGAGGAGATGAATCTAGATCTG 229055  
  
Qy 541 ggccttgataagcgcaccattgagcactacgttgacctgcagtagctgcccgagccagat 600  
|||||  
Db 229056 GGCTTGTATAAGCGCACCATTTGAGCAGTACGTGGACTTTGCACTAGCTGCCGAGCCAGAT 229115  
  
Qy 601 acccttcacgcgcagatttcccgcgttggagtcagctgcacccgcgaacagttcgtccgggc 660  
|||||  
Db 229116 ACCCTTACGCGCAGATTTCCGCTTGGAGTCAGCTGCACCGCAACAGTTCTGTCGCGGC 229175  
  
Qy 661 ggcagctggaaacgaagcgttacttcaagcctcagtcagtcacagatgagagtcgaaag 720  
|||||  
Db 229176 GGCAGGCTGGAAACGAGAGCGTTACTTCAAGGCTCAGTTCCAGAGTACAGAAGGTCGTAAG 229235  
  
Qy 721 ggtgaaggcagagacctcttcgacgcatcccgaggtgttggagatagcgtcgaaag 780  
|||||  
Db 229236 GGTGAAGGAGCAGGACCTCTTCGATCGCATTCGCCAGTGTTCGCGGGCATTTGACTCAACCGCA 229295  
  
Qy 781 catatgcgtgcgcagctgaccgtagctcgttcccttccgcggcagcttgcacttcaaccgca 840  
|||||  
Db 229296 CATATGCGTGGCGAGCTGACCGTAGGCTCGTTCCTTTCCGCGGGCATTTGACTCAACCGCA 229355  
  
Qy 841 attgcgcgcgttgaaaagcgcacaacccctgactgctcaacttcaaccacgggttccag 900  
|||||  
Db 229356 ATTGCGGCGCTTGCNAAGCGGCACAACCCCTGACCTGCTCAGCTTCCACACCGGTTTCAG 229415  
  
Qy 901 cgtgaaggctactcggaggtcgatgtgctgcgagtcgcgcgtcgagtgagcgtgag 960  
|||||  
Db 229416 CGTGAAGGCTTACTCGGAGGTGATGTGGCTGCGGAGTCCCGCGCTGCCATTTGGCGCTGAG 229475  
  
Qy 961 cacatcgtgaagattgtctcgcctgaggaataacgccaacgcgattccttaagatcatgtgg 1020  
|||||  
Db 229476 CACATCGTGAAGATTGTCTCGCTGAGGAATAGCGCAACGGGATTCCTTAAGATCATGTGG 229535  
  
Qy 1021 tacttggatgatactcgtagctgaccccatcatttggctccgcgtgtacttctgtggcagcgaa 1080





|    |      |                              |  |                                |                   |      |
|----|------|------------------------------|--|--------------------------------|-------------------|------|
| Qy | 601  | acccctcaacg                  | cgcagatttcccgccttgag                       | ctagagctgacgcgaacagttcgtccgggc | 660               |      |
| Db | 620  | ACCTTTCACGCGCAGATTTC         | CCGCGCTTGGAGTCAGGCTGCACCGAACAGTTTCGTCGCGGC | 679                            |                   |      |
|    |      |                              |  |                                |                   |      |
| Qy | 661  | ggcaactggaac                 | agaagcgttacttcaagcctcaagt                  | lccacagtacagaagctgctaaag       | 720               |      |
| Db | 680  | GGCAAGCTGGAACAGAAAGCGTTAC    | TTCAAGCCTCAGTTCC                           | CAGTACAGAAAGTTCGTAAAG          | 739               |      |
|    |      |                              |  |                                |                   |      |
| Qy | 721  | ggtaaaggacgag                | accctcttcgatcgcatgcccagg                   | tgttggaggatagcgtcgaaag         | 780               |      |
| Db | 740  | GGTAAGGACGAGGACCTCTTCGAT     | CGCATTCGCATTCGCCAGGTGTTG                   | GAGGATAGCGTCGAAAG              | 799               |      |
|    |      |                              |  |                                |                   |      |
| Qy | 781  | catatgcgtgcga                | agtgacctgacctagcttcgtttcctt                | ttccggcgccattgactcaaccgca      | 840               |      |
| Db | 800  | CATATCGCTCGCGACGTGACCGT      | TAGGCTCGTTCCTTTTCGGCGGCATTC                | GACTCAACCGCA                   | 859               |      |
|    |      |                              |  |                                |                   |      |
| Qy | 841  | attgcgcccttgc                | aaaagcgcacacacctgaactgcct                  | tcacctcaccacgggttctcag         | 900               |      |
| Db | 860  | ATTTCGGCGCTTGC               | AAAGCGCACAAACCTGACCTTG                     | CTCATTCAACACCGGTTTCGAG         | 919               |      |
|    |      |                              |  |                                |                   |      |
| Qy | 901  | cgtgaaggctact                | cggagtcgatctggtcgcggagtcgcgcgt             | cgatggcgtgagctgag              | 960               |      |
| Db | 920  | CGTGAAGGCTACTCGGAGT          | TCGATTCGCTCGGAGTCCGCGCT                    | CGGATTGGCGCTGAG                | 979               |      |
|    |      |                              |  |                                |                   |      |
| Qy | 961  | cacatcgtgaa                  | gatctcgcctgaggaatacgccaa                   | cgcgatctctaagatactgtg          | 1020              |      |
| Db | 980  | CACATCGTGAAGATTGCTCGCCT      | GAGGAATACGCCAACGGGATTCCT                   | TAAGATCATGTGG                  | 1039              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1021 | tacttgatgata                 | ctctgtagtaccatcattggtcccgcgt               | gtacttcgtggcagcgaa             | 1080              |      |
| Db | 1040 | TACTTGGATGATCCGTAGCTG        | ACCCATCATTTGTTCCCGCTGTACT                  | TGTGTTGTTGGACGGCGAA            | 1099              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1081 | gscgttaa                     | gcgctcaaggttgctgtctgagcgagcg               | cagatgagctgttcggtgga           | 1140              |      |
| Db | 1100 | GCACGTAAGCAGCTCAAGGTTG       | TGTGTCTGSGCAGGCGCAGATGAG                   | TGTTCGTTGTTGTTGGA              | 1159              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1141 | tacacatttaca                 | agagcgcgctatcgtgtcctcattg                  | agaagatacccttcccacta           | 1200              |      |
| Db | 1160 | TACACCATTTACAAGAGCGCGTAT     | CGCTGCTCCTCATTTGAGAAGATTCCT                | TCCCCACTA                      | 1219              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1201 | cgtaaaggcctg                 | ggaaaagctcagcaagttcttc                     | ccagacgcatgaaggccaagtcctt      | 1260              |      |
| Db | 1220 | CGTAAGGCGCTGGAAAGCTCAG       | CAAGGTTCTGCCACGACGCGCAT                    | TGAGGGCAAGTCGCTT               | 1279              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1261 | cttbgcgttgct                 | tcocatgacctggaagcgtacta                    | tcggcaacgctcgtctctcaat         | 1320              |      |
| Db | 1280 | CTTGACGCTGGCTTCCATGACCAT     | TGAAGAGCGCTACTACGGCAAC                     | CGCTCGCTCTTCAAT                | 1339              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1321 | ttcgaacagat                  | caacgcttatccatgggcaaa                      | gcggaatgggaccaccgcgaatc        | 1380              |      |
| Db | 1340 | TTCGACGATGCAACGCGGTATTC      | CATGGGCAAGCGCAATGGGAC                      | CACCGGAAAGTC                   | 1399              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1381 | actgcaccgat                  | cttaagcaaa                                 | ctccgcgaactttgatccagtag        | cccgcattgaacacctg | 1440 |
| Db | 1400 | ACTGCGCCGATCTACGACAGT        | CCCGCAACTTTGATCAGT                         | TAGCCCGCATGCAACACCTG           | 1459              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1441 | gactgttca                    | cttgatgctgcgcgcgacatcct                    | gtgtcaaggctgacaagatacaat       | catggcg           | 1500 |
| Db | 1460 | GATCTTTTCACTTGGATGCGCGCGACAT | CTCTGTGTCAAGGCTGCAAGAT                     | TCAGAGTTCAGAGACC               | 1519              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1501 | aactcccttg                   | agctcgagttccattcttgata                     | aggaagttttcaagggttgagagacc     | 1560              |      |
| Db | 1520 | AACTCCTTTGAGCTGCGAGTTCCAT    | TCTTGTGATAGGAAGTTTTTCA                     | AGGTTGCAAGACC                  | 1579              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1561 | attccttaca                   | lctgaagattgccaa                            | cgggtaccaccaagtacgcgt          | gcgcaggggaactc    | 1620 |
| Db | 1580 | ATTCTTACGACCTGNAGATTGCC      | AACCGGTACCAACCAAGTACG                      | CGCTGCGCAGGGCACTC              | 1639              |      |
|    |      |                              |  |                                |                   |      |
| Qy | 1621 | gagcagattgt                  | tccgcctcaagttttg                           | caccgcaagaagcttgggcttccctgtt   | ccccatg           | 1680 |
| Db | 1640 | GAGCAGATTGTTCCGCTCACGCTTTT   | TGACCGCAAGAAAGCTGGGCTTC                    | CGTTGTTCCCATGTTCCCATG          | 1699              |      |
|    |      |                              |  |                                |                   |      |

|    |      |   |      |
|----|------|---|------|
| Qy | 1681 | cgccactggcttgcggcgcatgagctgttcggttgggcgagagacaccattaaaggaaatcc  | 1740 |
| Db | 1700 | CGCCACATGGCTTGGCGGCGATGAGCTGTCGGTTGGCGCGAGGACACCATCATCAAGGAATCC | 1759 |
| Qy | 1741 | ggtactgaagatacttcaacaagcagctgtgctgatatgctgaacgagcagccgcgat      | 1800 |
| Db | 1760 | GGTACTGAAGATATCTTCACACAGCAGCTGTCTGGATATGCTGAACGAGACCCGCGAT      | 1819 |
| Qy | 1801 | ggcgtgcagatcatctcccctgcactgtgagctgttctgtcatttggctggcacggc       | 1860 |
| Db | 1820 | GGCGTGTCAAGATCATTCCTCCGTCGACTGTGCACTGTCTGTCATTTATGTGTGGCACGGC   | 1879 |
| Qy | 1861 | atttttgtgaaaaaccgattgatccacagatgagagaccgctccctaccggctcgagctt    | 1920 |
| Db | 1880 | ATTTTGTGGAACACCCGATGATTCACAGATTCAGACCGCTCTTACCACGTCGAGCTT       | 1939 |

RESULT

5

AE007071

LOCUS

DEFINITION

Mycobacterium tuberculosis CDC1551, linear BCT 27-APR-2001 complete genome.

ACCESSION

AE007071 AE000516

KEYWORDS

AE007071.1 GI:13881935

SOURCE

Mycobacterium tuberculosis CDC1551.

Mycobacterium tuberculosis CDC1551

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium tuberculosis complex.

1 (bases 1 to 14738)

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.

Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains

Unpublished

2 (bases 1 to 14738)

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.

Direct Submission

Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source

1..14738

/organism="Mycobacterium tuberculosis CDC1551"

/strain="CDC1551"

/db\_xref="taxon:83331"

/note="clinical strain"

complement(95..373)

/gene="Mt2246"

complement(95..373)

/gene="Mt2246"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAK46532.1"

/db\_xref="GI:13881936"

/translation="MSQARNPDPMARSGPDTSGGTAGPGAPVVIDCDCAARGSGCRDCVVSILGIVPELSLSDHERAALEVLADVLAPRLRIVPVRQRGSGVA"

480..2393

/gene="Mt2247"

480..2393

/gene="Mt2247"

/note="identified by match to TIGR protein family HMM TIGR00573"

/codon\_start=1

```
/transl_table=11
/product="DNA polymerase III, epsilon subunit, putative"
/protein_id="AAK46533.1"
/db_xref="GI:13881937"
/translation="MGATGGTQSLFADLAHAQGAATPADEMSLRRTFFVVVDLETTG
GRTTNDATPPDAITEIGAVKVCAGGAFELVNPQHSIPPQIVRLTGITAMVGN
APTIDAVLPMEFEAGDSYLVAHNAGFDIGFAAARCCDITWPOQVLCITRLARRV
LSDRAPSVRLAALAFVANSNTHRALDDARATVDVLIHALIERVNGVGYTAE LR
SYLNPVTOQRCKRVLAE TLPHRPGYILFRPGSGEVLYVGTAA DLRRLRVSOYFNCTDR
RRKMTMYLASSIDHVECAHPLEAGVRELMLSTHAPPYNRSKPEPYEMWVALTDE
APRLSVIRAPHRDRVGFRRSRKAAETAALLACTGLRTCTTRLTSARHGACPE
LEVSAPARVDTAQYAEAVLRAALIGGLDNAALAAVQVQVTELAERRRYESAARL
RDLTATLQALWHGQRLAALPELIAAKPDGPREGGYQLAVIRHQQLAAGRAPRG
VPMVPAVDIRGAQIILPTPAPLGALVEETALIRAWLAEPGRIVGVSYNDAAGLAS
PVRSAGPWAATAARSQALAGEQLSRGWQSDLPTEPHPSRQLFGRTGVDCTGPPQ
PLLPGRQPESTAG"
complement(2268..3311)
/gene="MT2248"
CDS
complement(2268..3311)
/gene="MT2248"
/notes="similar to PID:1743888; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="anthranilate phosphoribosyltransferase"
/protein_id="AAK46534.1"
/db_xref="GI:13881938"
/translation="WPSNPQILGRITDNRDLARGQAANWMDQIMTGNARPAQIAFAV
ANTMAKTADEVGEAGLVMLSHAPLPADTVDDADVVGVTGGDVNTVNLSTMAIV
VAAAGVPVYVGNRAASSGGADTLEALGVRI DILGPDIVARSLAEVGGFCFAPRFH
PSYHAAAVREIGTYPTVFNLPLTNPRPRAGLTGCAADLAELVMAGVFAARRSSV
LVHGDGDLDELTPTTSTIRVAAAGSVDKLTFDPAGFGFAQAQIDOLAGGDAQANAA
AVRAVLGGARGPVRDVAVLNAAIGAIVAHAGLSSRAEWLPAAWEEGLRRASAI DTGAAE
OLLARWRPGRQI"
3338..4149
/gene="MT2249"
CDS
3338..4149
/gene="MT2249"
/notes="similar to GB:D16254 PID:457676 SP:P50677;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cytochrome c oxidase, subunit III"
/protein_id="AAK46535.1"
/db_xref="GI:13881939"
/translation="WTSAVGTSGTALTSRVSLNRPNMVSGTIVWLSELMEFFACLF
AYFARAQAGNPPPTPELNLQAVPTVLIIASSFTCQMGVFAAERGDIFGLRRW
YVITLMLGLFFVGOAYEYRNLMSHGTSPSSAYGSEVFLATGFGHLHVTGGLIAFIF
LLVRTGMSKFTPAQATASIVSYWHEVDIVNIALFTVIYFIR"
4190..5032
/gene="MT2250"
CDS
4190..5032
/gene="MT2250"
/notes="similar to GP:3080504; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cytochrome c family protein"
/protein_id="AAK46536.1"
/db_xref="GI:13881940"
/translation="MTKLGFTRSGGSKGTRRRRLRRSLGGVLLLIATLIAGLAAV
LPTPQVAVADESSALLRTGKDFDTCVSGHANLQVDPHGP SLIGGGAAYVYFQ
VSTGMPARGEAQAPKRPDI FDEAQIDAIGAYVOAAGGPTVVRNPDGSIATQSLRG
NDLGRGDLFLNCASCNFTCKGGLSSGKYAPDLAPANEQIILTAMLTGQNPKPF
SNRQLSFEAKKDI IAYKVATEARQPGGYLLGGFGFAPEGMAWIIIGWVAALGLAWI
GARS"
5029..6318
/gene="MT2251"
CDS
5029..6318
/gene="MT2251"
/notes="similar to GP:4539218; identified by sequence
similarity; putative"
/codon_start=1
```

```
/transl_table=11
/product="iron-sulfur protein, putative"
/protein_id="AAK46537.1"
/db_xref="GI:13881941"
/translation="MSRADDADVGPPTCGGRSDDEERRIVPGPNPDQCAKDGAKATA
VPREDEAALAAASNOELLALGKGLDGVRIAYKEPRMPVEGTAKRAERSVAVWULL
GVGGLALILFI LFWPFEKAADGESDFTYSLTPLYGLTGLSLISIAIGAVLYOKR
FPIEETISIOERHDGASREIDRTKVANLTDAFEGSTIRRRKLIGLSFGVMGAFGLGT
LVAFAGLKNPKWPVPTAEGKKA VLTSGMTPRYOGETIYLARATGEDGPPFIKM
REDDMAGMETVFPWRESDGCTTVESHKLOETIANGIRNPVNLIRIKPSDLGRVVK
RKGDSFNGEGEFAFTKVCSHLGCPSSELYEQQSRYILCPCHQSQFDALHFAKPIFGPA
ARALAQLPITIDTGYLVANGDFVEVPFAPWERTTT"
6315..7964
/gene="MT2252"
CDS
6315..7964
/gene="MT2252"
/notes="similar to GP:4539217; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cytochrome b, putative"
/protein_id="AAK46538.1"
/db_xref="GI:13881942"
/translation="MSPKLSPPNIGEVLAQAEIDTRYHPSAALRRQLNKVPTHWS
FLIGETALYSFVVLITGYLTLFFDPDSMDVTYNGVQPLRGVEMSRAYQSALDISF
EVRGGLFVRQIIHWAALMFAAAIMVHLARFTGAPRRPRETNWVIGSLILLAMFEG
YEGSYLPDDLSCIGLRAALSSITLGMPIVCTLHWALFGDPGCTILIPRLALHIL
LIPGIIALIGLHALVWFQKHQTPGPGPQRTENHVGVMPVFAFKSAGFAAIVGV
LGLMGLLIQINPIWNLGPKYKSOVSAGSQDFYMMTEGLARIWPPWEYFWHHTIPA
PVMVAVIMGLVFLVPAYPLEKRETFDYAHNHLQRDPDRVPVRTAIGAMAIAYAVL
TLAAMDIIALFAHISLNTATWTRIGMIVLPFVYFITYRMCIGLQRSDRSVLEHGV
ETGIIKRLPHGAYIELQPLGVPVDEGHPILQYOGAPLPKRNMKLGSGAGSPGSGSFL
FADSAEADALREACHAASORALALREHQDSIMGSPDGEH"
complement(8084..8728)
/gene="MT2253"
CDS
complement(8084..8728)
/gene="MT2253"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46539.1"
/db_xref="GI:13881943"
/translation="WVSRYSAVRRGPDVISPVDIRILVGACAAVWLVTGVSAAAV
ALMDLGRGPHMAGNPHTTWLYAVIVVSALVIVGAIPVLLRARRNAEAPATRPITGA
SVRGSRISGSGHPAKRAVAESAPVQHADAFAEVAAEHSSSAVDRIHLRGTVLTSAGI
ALIAAATYILMAVGHDPGSWISYGLAGVVTAGMPVIEWLWIAQLARRVVAPOSS"
complement(8728..9627)
/gene="MT2254"
CDS
complement(8728..9627)
/gene="MT2254"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46540.1"
/db_xref="GI:13881944"
/translation="MSGPNPPGPEDEPESEPVSDTGRASGNHLLPPVAGGDKLPS
DOTGETDAYSRAYSAPESBHVTPPADRLIYDDIEESDLDDELAARWPMVY
GVAIIIAVALVVSLLVTRPHTSKLATGDTTSSAPVQDEITTTKPPPPPPAPP
PTTEIPATETQTVTVTPPPPPPTATTAAPPATTTAAAPPTTTTPTGPGQVTVSV
TGTKAPDIIISVTVYDAARPPRTQHNVIPWSMTVPTIPQSDVSGVSEASSLFRVSLN
CSITTSDCGVTLSSNSNDGPQTSC"
complement(9813..10232)
/gene="MT2255"
CDS
complement(9813..10232)
/gene="MT2255"
/notes="similar to GP:4539223; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical transmembrane protein"
/protein_id="AAK46541.1"
```







```
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 13157 AAGCGCGACAAGATAACGATGGCCAACTCCCTGGAGCTCGGGTGCCTTCTCGGACCG 13216
QY 1537 gaagttttcaaggttcagagaccattccttaagatctgaagatttcccaacggtaccacc 1596
Db 13217 GAGGTTTTTCGGGTGGCCCTCCCGGTTGCGGGCGGCGCAAGATCACCGGTACCACACC 13276
QY 1597 aagtcagcgtcgcagcagcactcagcagattgttcgcctcacattttgcaaccgaag 1656
Db 13277 AGTAGCGCTGGGCGCGGCTGGAGCCTATTGTCCCGCACACGCTGTGACCGGCC 13336
QY 1657 aagctgggtctcctgttcccatgcgccactggcttgcggcgatgagctgttcctggttg 1716
Db 13337 AAGCTCGGGTTCGGGTCCCGATCCCGATTCGGCTGCGTCCGGCGAGCTGTGGAGTGG 13396
QY 1717 ggcagagacacattaaagaatccggtactgaagatatttcaacaaagcaggtgtgctg 1776
Db 13397 GCGATGCGCAGCGGTGGGCTCGTCGAGGCGCGTCACTTGGTACATCGCCGCGGTAT 13456
QY 1777 gatctgtaagcagcagcagcagcgtgctgtgcagatcattcccgctgactgtgactgtt 1836
Db 13457 CGCATGCTCGACGAGCAGCGGTGGGCGCAGCAGCAGCAGCGCGGCTGTGGACCATG 13516
QY 1837 ctgtcatttatggttgcgcagcagcatttttggaaacgcgaattgatccacagattgag 1896
Db 13517 CTGATCTTTATGCTGTGGCAGCGGATCTCTGTCGAGCAGCAGCGTGTGCGCCAGATCAGC 13576
QY 1897 gaccgtctctaccggtcgagct 1919
Db 13577 GAGCCGACGATACCCCGTCCAGTT 13599

RESULT 7
MSGB1551CS/c
LOCUS MSGB1551CS 36548 bp DNA linear BCT 17-DEC-2001
DEFINITION Mycobacterium leprae cosmid B1551 DNA sequence.
ACCESSION L78813
VERSION L78813.1 GI:1377770
KEYWORDS
SOURCE Mycobacterium leprae (clone: cosmid B1551) (tissue library: Lorient 6) DNA
ORGANISM Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
1 (sites)
Eigmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae
Mol. Microbiol. 7 (2), 197-206 (1993)
93188700
2 (bases 1 to 36548)
Smith,D.R., Richter,P., Rubenfield,M., Rice,P.W., Butler,C.,
Lee,H.M., Kirst,S., Gunderson,K., Abendschan,K., Xu,Q., Chung,M.,
Deloughery,C., Aldredge,T., Maher,J., Lundstrom,R., Tulig,C.,
Falls,K., Imrich,J., Torrey,D., Engelstein,M., Breton,G., Madan,D.,
Nietupski,R., Seitz,B., Mao,J.I. et al.
Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome
Genome Res. 7 (8), 802-819 (1997)
97413161
9267804
This sequence was generated by the Genome Sequencing Center at
Genome Therapeutics Corporation (Collaborative Research Division),
100
Beaver St., Waltham, MA, 02154. Please contact Doug Smith
(smth@eric.com) for further information. The sequence represents
the
insert of a Lorient 6 cosmid clone from a mapped set of clones
constructed from M. leprae genomic DNA isolated from armadillo
liver
[3]. The sequence may not represent the entire cloned insert of
the
cosmid if an overlapping region was previously sequenced from
```

```
another clone. Coding sequences larger than 60 amino acids were predicted
on
the basis of codon usage and homology information. An attempt was
made
to locate the most probable start site based on codon usage,
homology,
the presence of a Shine-Dalgarno sequence, or overlapping orf that
suggested translational coupling. It is possible that the actual
start
site differs from the one selected.
FEATURES
Location/Qualifiers
1..36548
/organism="Mycobacterium leprae"
/specific_host="Dasypus novemcinctus"
/db_xref="taxon:1769"
/clone="cosmid B1551"
/issue_lib="Lorient 6"
/note="The liver of the armadillo was used to isolate the
Mycobacterium leprae."
BASE COUNT 6874 a 10335 c 11630 g 7709 t
ORIGIN
Query Match 37.3%; Score 716.8; DB 1; Length 36548;
Best Local Similarity 63.3%; Pred. No. 1.2e-160;
Matches 1134; Conservative 0; Mismatches 652; Indels 6; Gaps 2;
QY 133 gegttygattcaacgcgtctctccatcattgattgacacactcccacacacactgcgt 192
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11406 GCGTTTGGGTTTAAACCGGTTGTCCATCATCGACATTTCGCATTCGCATCGCGCTCGG 11347
QY 193 tggggacctgcggtgaacccgaccctcagcaactttcaacgctgagatctacaac 252
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11346 TGGGGACCAACCGGAGACACCAACCGCTACTCTGTTATTCAACGCGAGATCTACAC 11287
QY 253 tacgttgagctgcgttaagagc---tctcgatttgggatataccttaactctctggc 309
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11286 TACGTGAGCTGCGCAGCAGCTGGTCTATCCGCGACGGCGCTGCTTTCGCCACCGACGGA 11227
QY 310 gatggcagaccaattgttgcgtttccacacactggcgagtcctggttcgagcatctc 369
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11226 GACGGTGAATCTATCTGCGCGGCTTCCATTACTTGGGCGACCGAGATPACTAACACGCTG 11167
QY 370 cgcggaatttcgcgattgccatttgggatacaaaagaaagtcgctttctctgcgcgt 429
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11166 CGTGGCATGTTCCGCTTCGCGCTGTGGGACACCATGCCACAGAAATTGTTCTGTGCCCGA 11107
QY 430 gatcgttcggcatcaagccactgttctacgcaaccacccagcagcattggcaccgtttctcc 489
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11106 GACCCCTTCGGGATCAAAACCACTGTTTCATGCGCAGCGGAACCAACGCGCGGTGGCC 11047
QY 490 tcagagaagaagaccattcttggagatggccgagagagatgaatctagatctgggacctgat 549
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11046 AGCGAAAAGAAATTCCTTGCCTTGAATGCGCGAAGTATGATCTCGACACGAGATTGAC 10987
QY 550 aagcgcacattgacactcagtcgacctgacgtgcccgagccagcagatacccttcaac 609
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 10986 ACCCGGCACATACAGCACTACACGGTCTCTGAGTACGTGCGCGCGCCGAGACGCTGCAT 10927
QY 610 gcgcagatttcccgctggagtcaggctgcacccgcaacagttcgtccggcgcg---caag 666
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 10926 CGCGGGGTGCACCGACTCGAATCTGGCTGTTACGACGATTCGCTCCCGAACAGCTCGAT 10867
QY 667 ctggaacagaagcgttacttcaagcctcagttccccagtacagaaggtcgttaaggggtaag 726
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 10866 CCAGCGACCAACCGCTACTTTCGTGCCGAATTTCCGCCACACCGGATCACACGAGCACT 10807
QY 727 gacgagacctcttcgatcgcattgccaggtgttgaggaagcgtcgcgaagacatg 786
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 10806 GAACAGACCCGCTACGCGCGAATTCACCGCAGTCTTCGGGACTCAGTGGCGCAGACATG 10747
QY 787 cgtgcgcagctgacgctaggctcgttctcttcgcgcgccattgactcaacacgcgaattgcg 846
```



```
Db 10746 CGCGCCGACGTACCGTGGCTCGTTCTGTCCGCGGTATCGACTCCACAGCCATCGCA 10687
Qy 847 ccgcttgcaagcgcacaaacccctgacccgtcactcaccaccggtttcgagcgtgaa 906
Db 10686 GCATCGCGGATCCGACACAAATCCCGCGGTGATATACATTTACCAACCGGCTCGAGCATGAG 10627
Qy 907 ggcactcggaggtcgatgtgctgcgagtcgcccgctgcgattggcgtcgagcacatc 966
Db 10626 GGATCTCCGAGATCGACGTGGCGGTCTCTCGGCTGAAGCAATCGGTGCGCGCCACATC 10567
Qy 967 gtgaagattctcgcctgaggaatacaccacgcgattcctcaagatcatgtgtacttg 1026
Db 10566 GTGAAGGTGGTTACCCCAACAGAGTTCGTCGCCGCCCTGCCGGAGATCGTGTGATATCTC 10507
Qy 1027 gatgatctgtagctgacccatattgggtcccgctgtacttcgtggaagcggaaagcaagt 1086
Db 10506 GACGAGCCAATTTGCCAGCCCGCCCTGGGTTCGCTGTGTTTTTCGTCGCCGGAAGCTCGC 10447
Qy 1087 aagcagctcaaggttgctgtctggcgagggcgagatgagctgttcggtggtatacacc 1146
Db 10446 AAGCACGTCAAGTGGTGTCTCGCGGAGGGCGCGACGAGCTGTTCGCGCGCTACAGA 10387
Qy 1147 attacaagagcgcgtatcctgtgctccatttgagaagatcccttccccactacgtaaa 1206
Db 10386 ATCTATCGAGAGCGCTGTCTGTGAAGCCCTTCGACTACCTACCGGGGCCACTCGACGCG 10327
Qy 1207 ggcctggaagctcagcaaggttctgccagacgcatgcaaggcaagtcctctcttgag 1266
Db 10326 TCGATGGGCAAGGTGTCAAATCTCTGTGCCCAAGGTATGCGCGCAAGAATCTCTGTACAC 10267
Qy 1267 cgtggctcatgacatggaagcgtctactacgcaacgctcgtctctcctcaatttcgag 1326
Db 10266 CGCGGATCACTCACGCTCGAGAGCGTACTACGGCAATGCCGCAATTTCTCGGAAGCG 10207
Qy 1327 cagatgcaacgcgttattccatgggcaaacgcgaaatgggaccacgcggaagtcactgca 1386
Db 10206 CAGTTGCAGGCTGTGCTACCCAGGTTCCACACAGGCGCTGGACCCACACCGACGTCACGGCA 10147
Qy 1387 ccgactctacacaaatccgcaacttgatccagtacgtagccgcatgcaacacctgagctg 1446
Db 10146 CGGTGTACCGCGAATCTCGTGGGTGGGACCGGTAGCCCGCATGCGACATCGATCTG 10087
Qy 1447 ttcacctggatgcgcgacatcctgtcaaggctgacaagatcaacatgcccgaactcc 1506
Db 10086 TTCACCTGCCTACGTGGGACATTTTGSCCAAGGCCGACAGATGACAATGGCCAAATTCG 10027
Qy 1507 cttgagctcgagttccattcttgataaaggaagttttcaagggttgagagaccattcct 1566
Db 10026 CTGAGAGTCGCGGTACCGTTCCTTGATCCCGAAGTGTTCGCGGTGGCGTCCCGGTACCC 9967
Qy 1567 tacgatctgaagattgccaaacggtaccaccaagtacgcgctgcgacgggcactcgagcag 1626
Db 9966 ATGCAGGAAAGATCACCCGACCAACACAGAAATACGCATCTGCGGCGGCACACTGGAATCC 9907
Qy 1627 atgttcgcctcacggtttttgacccgcaagagctgggcttccctgttcccatcgccac 1686
Db 9906 ATCGTCCCGGCACACGTTGTCATCGCCCCCAACTCGGGTTTCGCGGTCCCAATCCGGCAT 9847
Qy 1687 tggcttcgscgatgagctgttcggttggcgagcagacaccattaaaggaatcgcggtact 1746
Db 9846 TGGCTGCGCGCGGGAAGTCTGGAATGGCGGTACGCGGTGGTGAATCTGCTCGAGGCC 9787
Qy 1747 gaagatacttccaacagcagcgtgtgctggaatgtatgtgaacgagcaccgcgagtggtg 1806
Db 9786 GAACATCTGGTCAACCTAGCCCGCTGCACCGGATGCTTGACGAGACCCGAGGTGGCATC 9727
Qy 1807 tcagatcatcccgctcagctdgcagctgttctgtcattataggtgtggcagcggcattttt 1866
Db 9726 AGCGATCATAGTCGCGCGCTGTGGACGGTGTAAATTTTATGCTGTGGCAGCGCATCTTC 9667
Qy 1867 gtgaaaaccgcattgatccacacatgagcagcgtcctcctaccgcgtcgagc 1918
```

Db 9666 ATCGAACACAGTGTGGTCCCGACAGATCAGCGAGCCACAGTATCCGGTACAGC 9615

RESULT 8  
MSGBI554CS/ c 36548 bp DNA linear BCT 17-DEC-2001  
LOCUS Mycobacterium leprae cosmid B1554 DNA sequence.  
DEFINITION  
ACCESSION L78814  
VERSION L78814.1 GI:1377771  
KEYWORDS  
SOURCE Mycobacterium leprae (clone: cosmid B1554) (tissue library: Lorist 6) DNA.  
ORGANISM Mycobacterium leprae  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium.  
REFERENCE 1 (sites)  
AUTHORS Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.  
TITLE Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae  
JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)  
MEDLINE 93188700  
AUTHORS Smith, D.R., Richterich, P., Rubenfield, M., Rice, P.W., Butler, C., Lee, H.M., Kirst, S., Gundersen, K., Abendschan, K., Xu, Q., Chung, M., Deloughery, C., Aldredge, T., Maher, J., Lundstrom, R., Tullig, C., Falls, K., Imrich, J., Torrey, D., Engelstein, M., Breton, G., Madan, D., Niatupski, R., Seitz, B., Mao, J.I. et al.  
TITLE Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome  
JOURNAL Genome Res. 7 (8), 802-819 (1997)  
MEDLINE 97413161  
PUBMED 9267804  
COMMENT This sequence was generated by the Genome Sequencing Center at Genome Therapeutics Corporation (Collaborative Research Division), Beaver St., Waltham, MA, 02154. Please contact Doug Smith (smith@eric.com) for further information. The sequence represents the insert of a Lorist 6 cosmid clone from a mapped set of clones constructed from M. leprae genomic DNA isolated from armadillo liver [3]. The sequence may not represent the entire cloned insert of the cosmid if an overlapping region was previously sequenced from another clone. Coding sequences larger than 60 amino acids were predicted on the basis of codon usage and homology information. An attempt was made to locate the most probable start site based on codon usage, homology, the presence of a Shine-Dalgarno sequence, or overlapping orf that suggested translational coupling. It is possible that the actual start site differs from the one selected.

FEATURES  
source Location/Qualifiers  
1. 36548  
/organism="Mycobacterium leprae"  
/specific\_host="Dasypus novemcinctus"  
/db\_xref="taxon:1769"  
/clone="cosmid B1554"  
/tissue\_lib="Lorist 6"  
/note="The liver of the armadillo was used to isolate the Mycobacterium leprae."  
BASE COUNT 6874 a 10335 c 11630 g 7709 t  
ORIGIN

Query Match 37.3%; Score 716.8; DB 1; Length 36548;  
Best Local Similarity 63.3%; Pred. No. 1.2e-160;  
Matches 1134; Conservative 0; Mismatches 652; Indels 6; Gaps 2;

Qy 133 gcgtttggattcaaccgcctctccatcatgtattgacactcccaaccactgcgt 192



|    |       |   |       |
|----|-------|---|-------|
| Db | 11406 | CGCTTGGGTTAAACCGGTTGTCATCATCATGACATATGGCATTCGCATACACCCGCTGCGG   | 11347 |
| Qy | 1193  | tggggaacctgcggatgaacccagccgcctacgcaatgactttcaacgggtgagatctacaac | 252   |
| Db | 11346 | TGGGACCACCGGAGACACCCAAACCGTACATCTGGTATTCAACGGCGAATCTACAC        | 11287 |
| Qy | 253   | tacgttagctgcgttaaaagc---tctcgagttgggataacatttaactcttgcgc        | 309   |
| Db | 11286 | TACCTGGAGCTGCGCGACGAGCTGTCATCCGGACACGGCGTGCTTTGCGCCACGACGGA     | 11227 |
| Qy | 310   | gatggcgaagcaattgttcgggtttccacacatcgggcgagtcctggttcgagcatctc     | 369   |
| Db | 11226 | GACGGTGAATCTATCGTCCCGGGCTTCCATTACTTGGGGCACCGAGATACTAACACGGCGTG  | 11167 |
| Qy | 370   | cgcgaaatgttcggcattgcgcatctgggatacaaaagaaagtcgctttcttcgctgcgt    | 429   |
| Db | 11166 | CGTGCAATGTTCCGCTTCGCGCTGGGACACCATCGCCCAAGAAATGTTCTGTGTGCCGA     | 11107 |
| Qy | 430   | gatcagttccggcataaagccactgtcttacgcaaacccagcgatgagcccggtttctctc   | 489   |
| Db | 11106 | GACCTTTTCGGGATCAAAACCACTGTTCATGTGCGCACCGGAACAAACGGCACCGCGTGCC   | 11047 |
| Qy | 490   | tcagagaagaagaccatcttgagatggcccgagagatgaatctagatctctgggcctgat    | 549   |
| Db | 11046 | ACGGAAGAATGCTCTTGAACTGCGCGAAGTGATCAGATTTCGACACGACGATTTGAC       | 10987 |
| Qy | 550   | aagcgcaacattgagcactacgtgaccctgacgtacgtgcccgcagcagatacccttcac    | 609   |
| Db | 10986 | ACCCGGCACTACAGCACTACACGGTCTGAGTACGTGCCCGACCGCCGACGCTGCAT        | 10927 |
| Qy | 610   | gcgcagattcccgcgttggagtcaggctgcaccgcacacagttctccggggcgg---caag   | 666   |
| Db | 10926 | CGCGGGTGCACCGACTCGAATCTGGCTGTTACGCACGCTATCCGTCCCGAACAGCTCGAT    | 10867 |
| Qy | 667   | ctggaacagaagcgttaacttaacgcctcaagtcctccagtcagaaggttcgtaaaaggtaa  | 726   |
| Db | 10866 | CCAGCGACCAACCCGCTACTCTGTGCGCGAATTTGCCGCCACACCGCATCCACGACGACT    | 10807 |
| Qy | 727   | gagcaggacctcttcgacgcattgcaccagtggttgaggagtagcgtcgaaaagcatatg    | 786   |
| Db | 10806 | GNACAGACCCGCTACGGCGGAATCACCGCACTGCTTGGGGACTCAGTGGCCACGACATG     | 10747 |
| Qy | 787   | cgtgcgcagctgacgtaggtcgtctcttccggcgccattgactcaacgcgcgaattgcg     | 846   |
| Db | 10746 | CGCGCCGACGTCACCGTTGGCTGCTTTCTGTCGGCGGTATCGACTCCACAGCCATCGCA     | 10687 |
| Qy | 847   | ccgcttgcaagcgccacaacccctgacctgtctacacttcacacacccggtttcgagcgtgaa | 906   |
| Db | 10686 | GCATGCGCGATCCGACACAATCCCGCGGTGATATCATTTACACCGGCTTCGAGCATGAG     | 10627 |
| Qy | 907   | ggctactcggaggtcgattgagctgcggagtcgcccgctgcgattggcgtcgacacatc     | 966   |
| Db | 10626 | GGATTCTCCGAGATCGACGTGGCGCTCTTTCGGCTGAAGCAATCGGTGCGCGCCACATC     | 10567 |
| Qy | 967   | gtgaagattctcgcctgaggaatacgcgaacgcgattcctaagatcatgttgactttg      | 1026  |
| Db | 10566 | GTGAAGGTGGTTACCCCCACAGAGTTCGTCCGCCGCTTCGCGGAGATCGTCTGGTATCTC    | 10507 |
| Qy | 1027  | gatgatcctgtagctgacccatcatattggtcccgctgtacttcgtgcgagcggaagcagt   | 1086  |
| Db | 10506 | GACGAGCCAAATTGCGAGCCCGGCCCTGGTTCCGCTGTTTTTTCGTGCGCGCGAAGCTCGC   | 10447 |
| Qy | 1087  | aagcacgtcaaggtttgctgtctctggcgagggcgagatgagctgttcggtgatacacc     | 1146  |
| Db | 10446 | AAGCACGTCAAGGTGGTGTGTGCGCGGAGGGCGCCGACGAGCTGTTCGGCGGGTACAGA     | 10387 |
| Qy | 1147  | atttacaagagcgcgctatcgtctgtctcatattgagaagatcccttcccactacgtaaa    | 1206  |
| Db | 10386 | ATCTATCGAGACCGCTGCTGTTGAGCCCTTCGACTACCTACCCGGGCCACTCGGACGC      | 10327 |
| Qy | 1207  | ggcctgggaagagctcagaaggttcttcgcagacgcgcatgaaggcgaaagtccttcttgag  | 1266  |

| Db         | 10326 | TCGATGGGCAAGGTGTCCAAATCTCTGCCCGGAAGGTATGCGCGGCAAGAAATCTGCTACAC  | 10267                  |
|------------|-------|---|------------------------|
| Qy         | 1267  | qgtggtccatgaccatggaagagcgctactacgagcaacgctcgctctcaatttcgag  | 1326                   |
| Db         | 10266 | CGCGATCACTCACGCTCGAGGAACGCTACTACGGCAATGCCGCAATTTCTCGGAAGCG  | 10207                  |
| Qy         | 1327  | cagatgcaacgcgttattccatgaggcaaacgcgcgaattgggaaccgcgcgaagtcaactgca  | 1386                   |
| Db         | 10206 | CAGTTGTCAGGCTGTGCTATCCACAGGTTCCACACGGCTGCAGCCACACACGACGCTACAGCA   | 10147                  |
| Qy         | 1387  | ccgatctacgcacaatccccgcgaactttgatccagtagccgcgcgcacactgagatctg  | 1446                   |
| Db         | 10146 | GGGTTGTACGGCAATCTGTCGGCTGGGACCCGGTAGCCCGCATGCAGCATCATCATCTG   | 10087                  |
| Qy         | 1447  | ttcaactggatgctgcgcgcgcacatcctggttcaaggctgacaagatcaacatggcgcaactcc   | 1506                   |
| Db         | 10086 | TTACCTTGGCTACGTGGCGACATTTTGCCAAAGCCGACAAAGATGACAAATGGCCAATTCG   | 10027                  |
| Qy         | 1507  | cttagctgagtgagttccattcttgataagaagttttcaagggttgcagagacaacttctc   | 1566                   |
| Db         | 10026 | CTGAGAGCTCGGGGTACCGTTCTTGTATCCCGAAGTGTTCGGCGTGGCGTTCGGGTACCC  | 9967                   |
| Qy         | 1567  | tacgatctgaagattgccaaagggtaccaccaagtagcgcctgcgcagggcactcgagcag   | 1626                   |
| Db         | 9966  | ATGCAAGCAAGATCACCCGACACACAGCAAAATACGCACCTGCGGGCGCATGGAATCC  | 9907                   |
| Qy         | 1627  | attgttcgcgcctcacggttttgcacgcgaagagctgggtcttcctctgttcccatgcccac  | 1686                   |
| Db         | 9906  | ATCGTCCCGGCACACGTTGATCGCCCCAAACTTCGGGTTCCCGGTCCCAATTCGGCAT  | 9847                   |
| Qy         | 1687  | tggttgcgcgcgcagctgctggttgggcgcaggacacacattaaaggaatccgggtact   | 1746                   |
| Db         | 9846  | TGGCTGCGCGCGCGGAATCTGCTGAATGGGGCTACGGCGTGGTGAACCTCGTCGCAAGGCC   | 9787                   |
| Qy         | 1747  | gaagatatctcaacaagcaggctgctggtatgctgaatgctgaacgagcagcgcgatggctg  | 1806                   |
| Db         | 9786  | GAACATCTGGTCAACCTAGCCGCCGTGCACCGGATGCTGTGACGACACCGAGGTGGCATC  | 9727                   |
| Qy         | 1807  | tcagatctccctgcagctggagctgtctgtcatattatggttggaacgagcgtatttt  | 1866                   |
| Db         | 9726  | ACCGATCATAGTCGCGCGGTGTGGACGGTGCTAATTTTATGCTGTGGCAGCCCATCTTC   | 9667                   |
| Qy         | 1867  | gtggaaaccgcgattgatccacagattgagaccgctcctaccgcggtcgagc  | 1918                   |
| Db         | 9666  | ATCGAACACAGTGTGGTGCCACAGATACGCGACGCCACACATATCCGGTACAGC  | 9615                   |
| RESULT     | 9     |   |                        |
| BSUB0016/c |       |   |                        |
| LOCUS      |       | 215640 bp   | DNA                    |
| DEFINITION |       | Bacillus subtilis complete genome (section 16 of 21): from 2997771 to 3213410.  | linear BCT 26-NOV-1997 |
| ACCESSION  |       | Z99119  |                        |
| VERSION    |       | Z99119.1  |                        |
| KEYWORDS   |       | GI:2635411  |                        |
| SOURCE     |       |   |                        |
| ORGANISM   |       | Bacillus subtilis.  |                        |
|            |       | Bacillus subtilis   |                        |
|            |       | Bacteria; Firmicutes  |                        |
|            |       | Bacillus/Clostridium group;   |                        |
|            |       | Bacillus/Staphylococcus group; Bacillus.  |                        |
| REFERENCE  |       | 1 (bases 1 to 215640)   |                        |
| AUTHORS    |       | Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borries, R., Boursier, L., Brans, A., Braun, M., Brignelli, S. C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmler, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghum, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Hatesch, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Huilo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaert-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., |                        |

Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Kurita,K., Lapidus,A.,  
Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,  
Masuda,S., Mael,C., Medique,C., Medina,N., Mellado,R.P.,  
Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,  
Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,  
Portetelle,B., Porwollik,S., Prescott,A.M., Prescan,E., Pujic,P.,  
Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,  
Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,  
Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,  
Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,  
Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,  
Takeuchi,M., Tanakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,  
Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,  
Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenecker,T.,  
Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,  
Yoshida,K., Yoshikawa,H.F., Zumsteln,E., Yoshikawa,H. and  
Danchin,A.  
The complete genome sequence of the gram-positive bacterium  
Bacillus subtilis  
Nature 390 (6657), 249-256 (1997)  
98044033  
2 (bases 1 to 215640)  
Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.  
Direct Submission  
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,  
Regulation de l' Expression Genetique, 28 rue du Docteur Roux, 75724  
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.  
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45  
68 89 48

FEATURES

source  
1. .215640  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
complement(8. .1001)  
/gene="ytnM"  
complement(8. .24)  
/gene="ytnM"  
complement(99. .1001)  
/gene="ytnM"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14906.1"  
/db\_xref="GI:2635412"  
/db\_xref="SPTREMBL:O34430"  
/translation="MKLIVFAFTGLLSQIDSLGMYGVTSLLLAFGITPAVAS  
ASVHLAEVPTAAGVSHIKFGNVKQTVQVLPQSIGAFGLAFLSQLPGDVAPY  
ISLFLLLGVLIRFLQYKPALEKRVPLNRKQSIPLGVIAGFADATGGGWPVT  
TPILLSRKUSPRKVVGVTVTSEFAIVASATAGFLISLGDVNWLMVFSLMAGGI  
APIAALVQKFLQMLGVGGFIILVNARTLINEWTANTAVHPLIYTAIGAIWLSAV  
LFVLSKIGNRINIKTAVDVHLKEK"  
complement(1020. .2270)  
/gene="hipO"  
complement(1020. .2270)  
/EC\_number="3.5.1.32"  
/codon\_start=1  
/transl\_table=11  
/product="hippurate hydrolase"  
/protein\_id="CAB14907.1"  
/db\_xref="GI:2635413"  
/db\_xref="SPTREMBL:O34980"  
/translation="MSLDYWRNTEGSPYOTTGNDILTLKEESNPVNLSTLEKOLIGI  
RRHLHOYPELSKEEFETAFIKKLEKGIQIRPTAKTKGVAFDIAGESPAIALRA  
DIDALPIEETKGLPYASKKHGKINHACGHDFHTAALGAFLKENQSLKGLIRLLQ  
PAEEAGATGYLIEDQGLDIDAVIGLHNKPDIAVGTGLKTPLMAAVDRFVETEG  
KGAHALPHNGFPDPIIGASQLIVALQITISRNYPQLQSAILTVKNGGSTWNIPTD  
VVEIGVTRTPDSVRNQKQRFVTEQISAAFLSKANVHKWSPPLCNDIAITGLV  
RDAHRAKQVLDIPAPSTAGEDFAFYLEHIPGSAFFGTGDDHWHHPAFTIDETAI  
KASYFLYESAKRLDLSNEESKISD"  
complement(2288. .2980)  
/gene="ribK"

CDS  
complement(2288. .2980)  
/gene="ribK"  
/function="regulation of riboflavin biosynthesis genes"  
/note="alternate gene name: ytnK"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="CAB14908.1"  
/db\_xref="GI:2635414"  
/db\_xref="SWISS-PROT:P94465"  
/translation="MTIAGTVVKGQKGRKLGFPPTANVDAKIHGLRNGVGVVATVN  
SLQHLGVMNIGVAPTVGNSLEKTELEIFDFHRDIYGEKIECSLFLKIRRERRDLSLE  
HQTKQIKKIDSCVAKREPLIGIMAPNKKESLLSHOELNLPDLACFYKCNINLYGNRVG  
YVNDNFFFEYGTQVAYRRIYILSFLSKEDNPVKVSSKYIRPGAGGLADKLNRFIS  
SYVEESENTLG"  
complement(3027. .4355)  
/gene="ytnJ"  
complement(3027. .4355)  
/gene="ytnJ"  
/function="unknown"  
/note="similar to nitrilotriacetate monoxygenase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14909.1"  
/db\_xref="GI:2635415"  
/db\_xref="SPTREMBL:O34974"  
/translation="MTRADFIQGAMIHGVGTTDGRWHPDVPDPSASTNIEFYMKKQ  
TAEKGLSEFIIADGLFISEKSIPEHLNREPITILSALASVTKNIGLVGTFSTPTE  
PFTISRLQMSLDHISGGAGNWLVSQEGAAARNHKSNNPEHTEREIAQERLDVVR  
GLNWSWEDAFIHNNKTGFDDQAKHLRHNKGYFOVEGNLPKQKQEPVVFQAG  
SSETGQFAAKNADAIPTHSNLEETKAFYADVKRAADEGRDPSVIRIFPGISPIVA  
DTEEAEEKYREFAELIPIENAVTYLARFFDDYDLVSYPLDEFPDIDGKNAFQST  
PTDKREAKARNLTREVAQEMAFPTLTFTGTPERVASLTETWFNAEAADGFIYGS  
DITLDAFVEXVITLQERGLYRQDYGRTGLRENGLGIPQHQSVLHSSH"  
complement(4352. .4633)  
/gene="ytnI"  
complement(4352. .4633)  
/gene="ytnI"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14910.1"  
/db\_xref="GI:2635416"  
/db\_xref="SPTREMBL:O34639"  
/translation="MSDVNVVMKSKGSCYCEVKYLNKEKGFPPQNDIVSEKEKLR  
DILQVKYGVHRVVPVVEIGRNOYQGITEIGIEHLDLALANHAQIKEAR"  
complement(4648. .5652)  
/gene="ytmO"  
complement(4648. .5652)  
/gene="ytmO"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14911.1"  
/db\_xref="GI:2635417"  
/db\_xref="SPTREMBL:O34846"  
/translation="MIRLSILDOSLIGEGETAADTLQHTVKLAQMAEECYGHRFWVAE  
HNNDETAGSAPVLLGYLASTRIRLASSGYMLOHYSYKYVAEOFHLLSALAPGRI  
DLVGKAPGGFQSLTDALQAEYKKPVQFDEKLEELTHFVRDDFPDTHRAALAPRQ  
VDRKPGIFLLGSTESAISAKLIGISFVFAFYINGEEVEVLKEARAEADHLPQGEAE  
FHLPAPFAATKEAEKHIYKRESIKVKLDGKRVNKGVSREQAEAYLENVTEPYDII  
VOKTGIITAGTKEEVAEBELTSLGTYKINDFVIPTPIKNAVEKQLSYLLSDAVLAAR  
"  
complement(5649. .6428)  
/gene="hisP"  
complement(5649. .6428)  
/gene="hisP"  
/codon\_start=1  
/transl\_table=11  
/product="histidine transport protein (ATP-binding  
protein)"  
/protein\_id="CAB14912.1"

gene

```
/db_xref="GI:2635418"
/db_xref="SPTREMBL:O34900"
/translation="MIEKNIHKQFIHHVLKGINLTVRKGBVVTIIGPSGSKTTEL
RCLNLEPDEGIIISIDKVINCRPESKKEVHLRKQAMVFOQVHLPAHKVIENVM
EGLTARKMRQDIAVAVENELRKVLQDKLNAYPSQLSGGOKQRVGILARALAIIPDV
LLEPTAARLDPELVGEVLEVMLEIATGKAMIVVTHMEFARRVSDQVVFMEGVIV
EOGTPEVFRHTKKDRTRQFLRRVSPLEYLFEPKEHIKEPVI"
gene
complement(6425..7132)
/feature="ytmM"
CDS
complement(6425..7132)
/feature="ytmM"
/function="unknown"
/feature="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB14913.1"
/db_xref="GI:2635419"
/db_xref="SPTREMBL:O34931"
/translation="WQDFPFIVSAMKEMVKPIPTLMMAVLPVIFGLVALIGNIIVR
IFRIKLVACSRFFYSFPTPAIHLIIMLYIGIPVADQVSSFHGLMSANEIIPVSM
FVIMALSTAGALYFEIIRSGILAMDTGOVEAAYSIGLTYSTFRVILPOLAKVSIIP
NFTNLIGLFLHTTSTAIVAVPEITGATIVASDNYAFLEAFIGAIIYVWLILLES
ANGVLERRAARFQGTSL"
gene
complement(7162..7881)
/feature="ytmL"
CDS
complement(7162..7881)
/feature="ytmL"
/function="unknown"
/feature="similar to amino acid ABC transporter (permease)"
/codon_start=1
Query Match 24.3%; Score 466.8; DB 1; Length 215640;
Best Local Similarity 55.7%; Pred. No. 7.5e-101;
Matches 997; Conservative 0; Mismatches 757; Indels 36; Gaps 4;
QY 88 caccgtggtctgcagatccggcaacttggcatgacgcgcgatgcagcttggattcaac 147
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128893 CACCGCGGTCCTGACAGTGATGATATTCATGATGAGCACGTCGCGTTCGGATTTCAGA 128834
QY 148 sgctctccatctgattgcacactcccaccactgcgttggggacctgcgcat 207
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128833 CGGCTCAGCATATTATGTAGAAAATGGTGACACCCCTTATCAT- - - - -AT 128786
QY 208 gaaccgcgcgtcacgaaagactttcaacggtgagatctacaactacactgcgtgagctgcgt 267
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128785 GAAGATGAACATATTGGATTATCTTTACGGAGAAATCTATAACTATATCGAAGCTGAGA 128726
268 aaagagctctcggatttgggatatacctttaatactcttgoggatggcgagccaattggtt 327
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128725 GAAGAACTTTGAAGCAAGGGGTACACATTCACACGCGATTTCGGACACAGAGGTTCTCTT 128666
QY 328 gtcggtttccaccacgggcaagtcggtggtcgcagcatctccgcggaatgttcgcgcat 387
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128665 GCGACTTACCGTCACATATTAAGAAAGACGCGCGCTCCAAACTGCGCGGTATGTTTGCCTTT 128606
QY 388 gccatttgggatatacaagaaagtcgtcttcttcgctgctgagatcagttcgcgcatcaag 447
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128605 CTGATTTGGAATAAAACGATCATGTGCTTTATGGGCAAGAGATCCATTCGCGCATTAA 128546
QY 448 ccaactgtcttaagcaaccaccagagatggaacggtgttctctccagagaagaccatc 507
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128545 CCGTTGTACTACACACAGCATCATATGATGATGAGTTTATTTTGCCTCAGAGAGAAAAGTTTA 128486
QY 508 ttggagatggcgcgagagatgaatctagatctggttcggttgataaagcgcaccattgagcac 567
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128485 ATGGTTGCTCAGAAATGATTAATA- - - - -GATAAAGAGCATTTGCAGCAG 128438
QY 568 tacgtggaactgcagtcagctgcgcgcgagcagatacccttcacgcgcgagatttcccgcttg 627
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128437 TACATGCTCTTCCAATTGTTCTGAAACAGCACGCTTGATGCCCATGTGAAAAAGTA 128378
QY 628 gagtcaggctgcaccgcgaacttcggttcggggaagctggaacgagagcgttacttc 687
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 128377 GAGCCCGGTTTACAATTCAAAATCCGCCCGCAGCGCATATCACATTTCAAAACGATATTC 128318
QY 688 aagcctcagttcccgatcacagaagcgtaaaagggtaagagcagagacccctcttcgatcgc 747
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128317 AAAGCGAACTTC- - - - -AAAGCCGTTTACAGTCAAGAACAGACTTGTGAAGAA 128267
QY 748 attgcccaggtgtttgaggatagctcgtaaaagcatatctcggtccgcgcgtagcgcgtagcc 807
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128266 GTCGAGAGACGCGATCTATGACTCTGTAAACGTTTCAATATGAGAAGTAGACGTTCTCTGTCGGT 128207
QY 808 tegtctcttccgcgcgcgattgactcaacgcgaattgcccgcgcttgcacaagcgcaccaaac 867
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128206 TCGTTTCTGTCTCGCGGATGATTTCTCTTTATTTGTGTCTGTTCGGAAGAAATTTTCAT 128147
QY 868 cctgacctgctcactccactccaccgcggtttcgcgcggtgaagcgtactcgcgaggtcgatgtg 927
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128146 CCAAGCTTAAACATCTCTGTGCGCTTTGAACAGCAGCGCTTCAGTGAGGTGTGATGTA 128087
QY 928 gctgcgcgagtcgcgcgcgctgcgattggtgcgtgcgcacatctgaagattgtctgcgcctgag 987
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128086 GCGAAAGAAACGCGCAGCGCGCTGGGCATTTGAAATATACGAAAGTCATTTGCGCTGAG 128027
QY 988 gaatagcgaacgcgcgattcctcaagatcatgtggtacttggatgatcctgtagctgaccca 1047
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128026 GAATACATGAATGAGCTTCCGAAGATCGTATGTCATTTTGATGACCGCTTGTGTATCCG 127967
QY 1048 tcatgtgtccgcgctgacttctgcgcgcggaagcgcgtaagcacgtcaaggttgtgtctg 1107
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127966 GCAGCGATTCCGTTGTACTTTGTGCAAAAGACGAAACATGTTACGGTTGCTTTTA 127907
QY 1108 tctggcgagggcgagatgagctgttctggtggtacacaccttaccaaagcgcgctatcg 1167
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127906 TCAGGTGAAGGAGCGGATGAGCTCTTCGGCGGATATAACATTTACCGTGAACCGCTTCG 127847
QY 1168 cttgctccatttgagaagatcccttccaccactacgttaaaagcctgggaagcgtcagcaag 1227
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127846 CTTTAAACCGTTTGAACGGATTTCGCTCGCGCTTGAAAAGATGCTTCTGCACGTTGCTGCC 127787
QY 1228 gttctccagacgcgcgcatgaagcaagtccttcttgagcgtggtcccatgacatggaag 1287
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127786 GTTATGCTTAAGGTATGAGAGGAAGAGCTGCTTGAGCGCGCTGCATCTCGCTTCAG 127727
QY 1288 gagcgtactacggcagcgcgtcgtctctcaatttgcgcgcagatgcacgcgtttattcca 1347
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127726 GATCGTTATATCGTAAACGCGAAAAATCTTTGAAGAGTCTGTCAAAAGACAGCTGTTCAAG 127667
QY 1348 tgggcaaaagcgcgaatggggaccaccgcgaagtcactgcacccgatctctacgcacaatccgc 1407
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127666 CATTAATAATCCGAACCTTTTCATATCGCGATGTGACGAAGACTTTACTTTACAGAAGCAGC 127607
QY 1408 aactttgatccagtagccgcgcgcgcacacctggtatctgttccactggatgcgcgcgcac 1467
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127606 TCGTACAGCGATATCAACAAAATGCAGTATGTCGATATCCACACTTGGATTCGCGCGCGAC 127547
QY 1468 atcctggtcaagcgtcacaagatcaacatggcgcaactcccttgagctgcgagttccattc 1527
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127546 ATTTTGTAAAGCGGATAAAATGACGATGCGGAAATCTCTTTTGAAGCTGCGTGCCTTC 127487
QY 1528 ttggataaggaagtcttcaaggttgagagagaccattcttcaatctgaagattgcaccaac 1587
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127486 TTGGATAAAGTCGTATTTGACGCTCGCTTCAAAAATTTCTCTGATGAGCTGAAAACGAAGAC 127427
QY 1588 ggtaccaccaagtcacgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1647
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127426 GGCACGACGAAGTATCTTCTTCGCAAAAGCCGCTGAAGGAATTTGCTCTGACGACGATATTA 127367
QY 1648 caccgcaagaagctgggcttccctgttcccatgcgcgcacctggcttgcgcgcgagtgagctg 1707
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127366 AACCGCAAAAAGCTTGGTTTCCCTGTACCGATCCGTCATTTGGCT- - -GAAAAACGAATG 127310
QY 1708 ttcggttgggcgcgagacaccataaggaatcccggtactgaagatatcttcaacaagcag 1767
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127309 AACGAATGGGTGCGGAATATTATTCAAGAAAGCAACGGATGCTTATATCCATAAGGAT 127250
```



operate in *Bacillus subtilis*: characterization of OpuD  
J. Bacteriol. 178 (17), 5071-5079 (1996)

96359357 MEDLINE  
22 (bases 1 to 220060) REFERENCE

22 (Pages 1 to 220000)  
 AUTHORS  
 TITLE  
 JOURNAL  
 Microbiology 143 (Pt 11), 3431-3441 (1997)  
 In the 200 kb *rrnB*-*dnaB* region  
 Sequencing and functional annotation of the *Bacillus subtilis* genes

98048467 MEDLINE  
23 (bases 1 to 220060) REFERENCE  
Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, D. AUTHORS  
Direct Submission TITLE  
Submitted (13-JUN-1997) Laboratoire de Genetique Micro  
Domaine de Vilvert, Jouy-en-Josas cedex 78352, France JOURNAL

Query Match 24.3%; Score 466.8; DB 1; Length 220060;  
Best Local Similarity 55.7%; Pred. No. 7.6e-101;  
Matches 997; Conservative 0; Mismatches 757; Indels 36; Gaps 4;

Qy 88 caccgtggtcctgacgatgccggcaacttggcatgacgccgatgcagcgtttggattcaac 147





|            |   |   |                            |
|------------|---|---|----------------------------|
| Qy         | 1768  | gctgtgctgatatctgtgaacagagacccgcgtagtgctgtagatcatctccgctgactg    | 1827                       |
|            |   |   |                            |
| Db         | 63362   | TACGTGTGACTAGACTTACTAGACGATCACTGTGCGCGTAAATTCGACTACGTCGTAATAATC | 63303                      |
| Qy         | 1828  | tgcactgtctctgcattatggtgtgcacgcgcatttttgcgaaaac                  | 1875                       |
|            |   |   |                            |
| Db         | 63302   | TGGACACTCGTATTTTCATGATTTGGTACGATGCTATACGTACGACGAC               | 63255                      |
| RESULT     | 12  |   |                            |
| LOCUS      | AL596169/c  | 239050 bp   | DNA linear BCT 04-DEC-2001 |
| DEFINITION | Listeria innocua Clp11262 complete genome, segment 7/12.  |   |                            |
| ACCESSION  | AL596169  | AL592022  |                            |
| VERSION    | AL596169.1  | GI:16414035   |                            |
| KEYWORDS   |   |   |                            |
| SOURCE     | Listeria innocua.   |   |                            |
| ORGANISM   | Listeria innocua  |   |                            |
| REFERENCE  | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacteriophages; Firmicutes; Bacillus/Staphylococcus group; Listeria.  |   |                            |
| AUTHORS    | 1 (sites)   |   |                            |
|            | Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P., Dommann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K. D., Fsihi, H., Portillo, F. G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F., Kurapkut, G., Madueno, E., Maistounam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordstiek, G., Novella, S., de Pablo, B., Perez-Diaz, J. C., Purcell, R., Remmel, B., Rose, M., Schluter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P. |   |                            |
| TITLE      | Comparative genomics of Listeria species  |   |                            |
| JOURNAL    | Science 294 (5543), 849-852 (2001)  |   |                            |
| MEDLINE    | 21537279  |   |                            |
| REFERENCE  | 2 (bases 1 to 239050)   |   |                            |
| AUTHORS    | Glaser, P., Frangeul, L. and Rusniok, C.  |   |                            |
| TITLE      | Direct Submission   |   |                            |
| JOURNAL    | Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 13, FRANCE   |   |                            |
| COMMENT    | E-mail: pglaser@pasteur.fr  |   |                            |
| FEATURES   | Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.   |   |                            |
| source     | Location/Qualifiers   |   |                            |
|            | 1. .239050  |   |                            |
|            | /organism="Listeria innocua"  |   |                            |
|            | /strain="Clp11262"  |   |                            |
|            | /db_xref="taxon:1642"   |   |                            |
| gene       | complement(57..1355)  |   |                            |
|            | /gene="lin1549"   |   |                            |
| CDS        | complement(57..1340)  |   |                            |
|            | /gene="lin1549"   |   |                            |
|            | /note="similar to unknown protein"  |   |                            |
|            | /codon_start=1  |   |                            |
|            | /transl_table=11  |   |                            |
|            | /protein_id="CAC96780.1"  |   |                            |
|            | /db_xref="GI:16414036"  |   |                            |
|            | /translation="MAIQPLAYMRPKALDEIVGTHLVGKDKIIRVMYKAKOLSSMI LYVPGIGKTSASATAGSTKTAFTRLNATNNKDMVEVAAEAKMSGTVILLDEIVH RLDKADKPLLPESGAILLIGATSNPYIAINPAIRSRTOIFELKPLTVEDIMITH DRALLDKPRLGNLVEIDEIDEPKAKHPATASGNDVRSALNALELAVISSPNEGDIITH TLDAVECEQLKSLANDKDGDAHYDLSAFQKSVRCSDVNAALHYMGRLEIAGDLVSI SRMLVNPDEIGLANPQAGHTLAAIQTAERKVGFPPEARIPLANAVDELCLSPKSNKSA IMAIDALADIROGNSGEVPHLDHGHYSGAKKLGKRAMDYKYPHANVDELANWDDQYLPD RLKMKLYDPKFTSKFEQTIAGYQIKINENAKKE"  |   |                            |
| RBS        | complement(1350..1355)  |   |                            |
|            | /gene="lin1549"   |   |                            |
|            | 1490..1938  |   |                            |
| gene       | /gene="lin1550"   |   |                            |
|            | 1490..1912  |   |                            |
| CDS        | /gene="lin1550"   |   |                            |
|            | /note="similar to unknown protein"  |   |                            |
|            | /codon_start=1  |   |                            |



```
/transl_table=11
/protein_id="CAC96781.1"
/db_xref="GI:16414037"
/translation="WKITKGRYGLTITILELAKRIGDGPISLRSIAQDNKLNSEHYLEQ
LIGPRNRAGVIRKISVYVNGDPKIKTAGDIIIRTEGPIVLVSEMEDEAAQRE
LWTRNRNAVRLDQTTLSDDLKHSTDSLTGMYFYI"
1921..1938
/gene="hiss"
/codon_start=1
/gene="lin1550"
2144..2149
/gene="lin1551"
2144..3362
/gene="lin1551"
2157..3362
/gene="lin1551"
/note="similar to ammonium transporter NrgA"
/codon_start=1
/transl_table=11
/protein_id="CAC96782.1"
/db_xref="GI:16414038"
/translation="MESVFMFFCTLLVLMTPGIALFYGMVRRKNVLSTAMYSFSM
AIIILWTVGSLAFAPGNGFIGSFDWTFLHNVGFAANDTYSDAIPHILPMFWOTF
AIIITVAIISGAFAERNPSAYLIFILMSLLVYSPVAHWGDDGKWLRELGDALDAGG
NVVHSSGVTGLVLAIMIGRRKADSASPHNPLALIGLILWFYGFNGFSALTIID
NVAMTAFVNTAAAGIIGWGLVEMLNKKPTMLGTMLGTSAGIAGLVSITPPAAGFTVP
SSLIIIGFLGGALCFWAWFLKGVKRYDDALDAFLHGIGGIWGGIATGLFATKYNEA
GADGLFYGNASLVKQLIAIGTSVAVVAVTALIVVVIKFLPIRVNREQYKGLDIT
LHGERAYOB"
3376..3741
/gene="lin1552"
3376..3741
/gene="lin1552"
/note="similar to nitrogen regulatory pii protein"
/codon_start=1
/transl_table=11
/protein_id="CAC96783.1"
/db_xref="GI:16414039"
/translation="MSGLTKEIIRPNRHFHFORLAKIGVSGLTVTVKALGTGLEKG
FIELRGYTKESNVHERMKIEIWTVPVEDVLRVVKETLRTGEPDGRKVFYPLAEV
VKISGTGIDALQDKPAK"
3872..3877
/gene="lin1553"
Join(3872..4271,4279..4298)
/gene="lin1553"
3891..4271
/gene="lin1553"
/codon_start=1
/transl_table=11
/protein_id="CAC96784.1"
/db_xref="GI:16414040"
/translation="MRKIPIFTLLAAILLFGGYLYVQHASSKVDVTNFVVDQDKNN
TVTKIDGAGNKSITPLPSKESMDTVKGNRRVNEATFYNNKISSSEAKELDGPFW
NDSNGLLVNKHVNIQELSDDF"
4279..4298
/gene="lin1553"
complement(4279..4298)
/gene="aspS"
complement(join(4279..4298,4307..6082))
/gene="aspS"
complement(4307..6082)
/gene="aspS"
/codon_start=1
/transl_table=11
/product="aspartyl-tRNA synthetase"
/protein_id="CAC96785.1"
/db_xref="GI:16414041"
/translation="MEKRTSYCGELNETHIGOSVILHGWVQRRLDGLGIFIDLRDR
GIVQVVPFPFIEDGVNVSDLELRKYRLDLRRPMMNFIKMRHTVTRFNKIDAL
TLNLTSTKTPFIEDGVNVSDLELRKYRLDLRRPMMNFIKMRHTVTRFNKIDAL
GFDEIETPLTSTEGARDYLPVSRYPGNYALPQSPQILKQLLMTAGFKYQIV
RCFREDRUGRQPFETQIDLETSLTKEEQIATEDMLVDVVVKEAKNITIEKPFPM
TYKEMDRFGSDPDIRFGLLEQNVSEVVKDYDFKVSAIENGGEVKAINKAATAFN
FSRKDLALGVFVANYGAKLWLKIEAGELGPIAKFFPEDKAAELQAALQAEQDL
LLFAADKADIVAASLGALRNKLGKLDLINEDELAFLWVTDWPLFEYDEACRGYSAH
```

```
HPFTLPKERDIPFLETSSKVMAEAYDIVLNGYEIGGSLRIYKKEVQSMFRALGFT
DESAKEQFGFLEALEYGPPhGGTALGLDRIVMILAGRNLRDITAFPKTCSAVDPL
TNAPQESAAQLAELKLETVKKEYN"
complement(6084..7361)
/gene="hiss"
complement(6084..7361)
/gene="hiss"
/codon_start=1
/transl_table=11
/product="histidyl-tRNA synthetase"
/protein_id="CAC96786.1"
/db_xref="GI:16414042"
/translation="MDLQLPRGTROIPEEVSKWHFLESFAFNKVCNYSQYEEIIRPIF
EHTELFERGVGDSITIVSKEMVTFIDKGRSLTLRPEGTASVVRAVEHKLKVEVSQP
IKWYNEMFVRPQGRQROFTOMGIEALGSDPSIDVEVTSLAMEFRKIGLTNI
KLVINSGLDKESRLKXREALVAHFPHIDEFCAECOVRLHKNPLRLDCKKDNDNPLI
QSAISLDFLNEESVAIFENKYLTALEIPFIDPTVRGGLDYIWNHTTFEIMSVEEG
FGAKTTLCCGGRYHGLVKEFGPDPPTGIGFVGIVERILLALEKAETNIPTKPLEVYV
ITAQPEALIKAVTLVNLKQNGISAEKDYLRKRLKAQLKDANRKAIIYTVILGEELQ
TGYOLKNMTEGEAVSETTIEKLTNTEEK"
Join(7803..9086,9092..9130)
/gene="lin1556"
7803..9086
/gene="lin1556"
/note="similar to N-acetylmuramoyl-L-alanine amidase"
/codon_start=1
/transl_table=11
/protein_id="CAC96787.1"
/db_xref="GI:16414043"
/translation="MKNKEFITVVSILLIAAGIFTTIANANANSVVVAEVLNVRSG
PGLAYDVTQARKNEVLRVGEENQWYKVLQDNGSGWASLWENTDVSAAANSIAI
VSDGGLNVREKPSSTSLGLNNGDQVTVTSQONGWAOIQYNGKSAVWSYOYLIR
ESVTKVDESELOTVTIRDDSTNIRNKPGRGAVIEKANSQGFAGIOGVODWYKIRTT
SSEGVYANWYVDVSDKQTSRPTTKLSEATIVIDPGHGNDDPGKANGATIEKE
MTLKTAKLUKLESGAKVILTRNSDKYSLKSRTNAAENKADVFISIHFDLSLEDS
KSVSQDTTYYYDNDSDKSLAESINTTLGNDLPTTNRGSRVGVYVVRNSQPAVLLEL
GYLSAKDRNINSASYSQIADSVTDLSNYSFN"
complement(join(9092..9130,9135..9587,9597..9602))
/gene="lin1557"
complement(9092..9130)
/gene="lin1557"
9092..9130
/gene="lin1556"
complement(9135..9587)
/gene="lin1557"
/note="similar to unknown proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC96788.1"
/db_xref="GI:16414044"
/translation="MRVLLQRCYEASVRVEGEIISIEAGLCLLVGTFHSDTEETVEY
MAKTIIGLVFEDESKMNLISLAERGAILLSVSQFTLYADVSRGKRPSPTKSPAGEKA
EGLYNLFNNKLSDAGFIVETGVGFAMFDKVIYVNGPVTIMLDSEMRK"
complement(9597..9602)
/gene="lin1557"
complement(9603..11834)
/gene="relA"
complement(9603..11819)
/gene="relA"
/note="similar to (p)ppGpp synthetase"
/codon_start=1
/transl_table=11
/protein_id="CAC96789.1"
/db_xref="GI:16414045"
/translation="MAKEQNLTAEQVIDMASHYMNQEHLLALVKKAYEFARDSHKEQFR
```

```
Query Match 20.1%; Score 386.6; DB 1; Length 239050;
Best Local Similarity 53.3%; Pred. No. 11e-81;
Matches 952; Conservative 0; Mismatches 799; Indels 36; Gaps 5;
QY 88 caccgtggtcctgacgatgcgcgaccttgccatgacgccgatgcagcgtttgattcaac 147
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 219817 CACCGTGGCCAGCATGATGAAGGATACTTCCACAGATCATCAGTCGCGATTTGGTTCCGC 219758
```



## JOURNAL

Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK Unitté de Genétique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk

Notes:  
Details of M. leprae sequencing at the Sanger Centre are available from [http://www.sanger.ac.uk/Projects/M\\_leprae/](http://www.sanger.ac.uk/Projects/M_leprae/) A relational database containing the M. leprae sequences is available from <http://genolist.pasteur.fr/Leprae/>.

## COMMENT

## FEATURES

source  
1. .348450  
/organism="Mycobacterium leprae"  
/strain="TN"  
/db\_xref="taxon:1769"  
551. .1786  
/gene="ML0842"  
551. .1786  
/gene="ML0842"  
/note="Similar to Mycobacterium tuberculosis hypothetical 44.6 kDa protein Rv1464 or Mtv007.11 TR:O53155 (EMBL:AL021184) (417 aa) fasta scores: E(): 0, 43.6% id in 408 aa and to many other bacterial nifs-homologues, e.g. Bacillus subtilis yurW protein yurW TR:O32164 (EMBL:Z99120) (406 aa) fasta scores: E(): 0, 46.7% id in 405 aa. Previously sequenced as TR:O32975 (EMBL:Z98741). Contains Pfam match to entry PF00266 aminotran\_5, Aminotransferases class-V. Similar to ML0117, ML0596 and ML1708"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAC31223.1"  
/db\_xref="GI:13092923"  
/translation="WDVFDIQAVRADPFIQETVNGKPLIWFNDNAATQKQVVIDR VSYFAHNSNTHRAAHELAARATDAYEARETARRFTGAAKAQEIIFVRGTAEINL VAYANGKHQPGDEVITLHEHIANIVPQLLSQGTGAILKAVPDDAGNLMSEFE DLLGPRTKLVAGTQVSNALGTQGEIKVELGHRYGARLIDGAOSIPLHPNINVELG ADFVFSGHKIVGPTGIVLYGCEVDLTMPQGGNNIVDVTLESLYOGPPNKFE AGTGNADALGELALRYVERVGVORIASHEQALDYATPRLADIPGVRVLTGATAEKA SVLSFVLAHGEFLEVGKALNAEGIAVRAGHCAQVPLRLGLEATVRPSFAYNTIYE IDVFNVVRIRIAEGSTNIG"  
653. .1738  
/gene="ML0842"  
/note="Pfam match to entry PF00266 aminotran\_5, Aminotransferases class-V, score 23.70, E-value 1.7e-08"  
1908. .2395  
/gene="lppp"  
/note="ML0843"  
1908. .2395  
/gene="ML0843"  
/note="Possible pseudogene of M. tuberculosis orthologue lppp (Best blastx score 294)"  
/codon\_start=1  
/pseudo  
/transl\_table=11  
/product="lipoprotein (pseudogene)"  
2549. .2555  
/note="possible RBS"  
2559. .4112  
/gene="nark"  
/note="ML0844"  
2559. .4112  
/gene="nark"  
/note="Similar to Mycobacterium tuberculosis putative nitrite extrusion protein narK1 or Rv2329c or MTCY3G12.05 TR:P71883 (EMBL:Z79702) (515 aa) fasta scores: E(): 0, 69.3% id in 488 aa and to Escherichia coli nitrite extrusion protein 2 narU SW:NARU\_ECOLI (P37758; P77696) (462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There is a frameshift near the C-terminus relative to the M. tuberculosis homologue. Previously sequenced as TR:O32974 (EMBL:Z98741). Contains hydrophobic, probable

gene  
CDS  
misc\_feature  
gene  
CDS  
RBS  
gene  
CDS

membrane-spanning regions."  
/codon\_start=1  
/transl\_table=11  
/product="putative nitrite extrusion protein"  
/protein\_id="CAC31225.1"  
/db\_xref="GI:13092924"  
/translation="MDQVLLQAEESLYGRRSKTVPIDHRHLSKSSFFRLRLRRAC RISHWDPQAAAEAGNKTIARLLNLSVTLVHLGYSVTLPVMELEFMPKDYGFSA GPFLLATATLVAGCLRPVSLATALFGRNWAIFSVMLIPIITATVVLAAHPGLP LAPYLACAATLGLGGNFAASMTNANAFPHRLKGAAGLAGAGNLGYVIOVCGML VIASVDKRPYIVCGLYVYVLLIIAGIGAMLFPHNDIEHHRIGYNTIRPVLVYVSTRDS WYLLALYLASFSGFIFGFAQGVLETFNFAAGQSTAQAALHAAELAFIGPTLAAVAR FWGRLADRLGGSRTLVYFAGVFAAGLGLVIGIIEGSRVCPIRGVMAASYPAGFIT LFTLSGLNGSVYKMIPTIFEACSHSLGINDDECRDMSRVISGVVIGFVAEVALGCV GIDLALRESYLVNTGGVTAAFWIFMLCYAAAAGVLTWKMYVCRPLPGNLHDEAANAFAAS VGASRTHRG"  
complement(4100. .4681)  
/gene="ML0845"  
complement(4100. .4681)  
/gene="ML0845"  
/note="Unknown function. Similar to part of some acyl-CoA oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2 TR:O65201 (EMBL:AF057043) (692 aa) fasta scores: E(): 1.5e-06, 35.2% id in 125 aa. Previously sequenced as TR:O32973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0, 99.5% id in 193 aa."  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAC31226.1"  
/db\_xref="GI:13092925"  
/translation="MPLRPTNAASERVTKRTATETILQRMVGARQSEKKSFLNRGT OGTIFEKCEDCLLSAVARRQVKSKEMFADFTNTVDHLLHAANSHTRVVLLETFLA GIESCNDPAARLLDIYCDLYALSKARHGTSSIDTSPFSALKAIAGSRPMPSAAAL CRDVGRRFSSESQSCVTPRCRTSYKRCPSQLQPR"  
complement(4959. .5480)  
/gene="ML0846"  
complement(4959. .5480)  
/gene="ML0846"  
/note="Possible pseudogene similar to M. tuberculosis paralogue Rv1747 (Best blastx score 127)"  
/codon\_start=1  
/pseudo  
/transl\_table=11  
/product="ABC transporter (pseudogene)"  
complement(5522. .5848)  
/gene="ML0847"  
complement(5522. .5848)  
/gene="ML0847"  
/note="Possible pseudogene of M. tuberculosis orthologue Rv2327 (Best blastx score 184)"  
/codon\_start=1  
/pseudo  
/transl\_table=11  
/product="conserved hypothetical protein (pseudogene)"  
5906. .8080  
/gene="ML0848"  
5906. .8080  
/gene="ML0848"  
/note="Similar to Mycobacterium tuberculosis hypothetical ABC transporter ATP-binding protein Rv2326c or MTCY3G12.08 SW:YN26\_MCTU (P71886) (697 aa) fasta scores: E(): 0, 76.9% id in 697 aa. Shares similar domains with many ABC-type transporters e.g. Streptomyces roseofulvus ATPase component of putative ABC transporter frnD TR:O68910 (EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5% id in 495 aa and Synecococcus sp. nitrate transport ATP-binding protein NrtD nrtD SW:NRTD\_SYNP7 (P38046) (274 aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa. Previously sequenced as TR:O32971 (EMBL:Z98741). Contains hydrophobic, possible membrane-spanning regions. Contains 2 Pfam matches to entry PF00005 ABC\_tran, ABC transporter. Contains 2 x PS00017 ATP/GTP-binding site motif A

```

(P-loop). Contains 2 x PS00211 ABC transporters family
signature."
/codon_start=1
/transl_table=11
/product="ABC transporter"
/protein_id="CAC31229.1"
/db_xref="GI:13092926"
/translation="MDICIHRRMALLCOATRRHRWKTLQRCRAAPDLAPGTGTIGKV
TAHDPRIHLHRSGLQPELAQASVLAALCAVTAIVSVVPPFAAGLALLIGTVPMGLLA
YRFBYRLMTAMWAAGVIAELIHLGIGFIAVNASAYIGLTVVYKRGQGLTVIALA
LFAGLAFGANVVALVVLGRHLIRFRAMTANVDGIAATLHMLHPVAOLKRYFAD
GLOHPWMLGYFVITILVSVLSIGWVLSRVLERIDIPDVHKLIDAPSAKNEDAPVG
PVMWLDKRYFPHAGDARLESWDLRVGEHVAVTGANGSKTLMILLAGREPTSG
TADRPAGVLGKGGTAVLQHPESVGLTRVADVVYGLPPGTDVDVNRLLRREVLGD
AFADRDTGSLSSGLOLALAAALAREPSLLIADENTSMVDROGRDALLGLVSLTKR
HPIALVHTIHNNEADRTINLSDSPDNAGMAETVAPVSTVAVDHRHPHVLVLE
GVGHEXSGSTPWAKAALHDISFVYRQGDGVLYVGSNGSGKSTLAWINAGLMVPTGAC
LIDGRHERHVAVALSFQARQLMRSDLEVASAGFSPRDEHRAALGVVVGUD
PALAKRIDQLSGQMFRVVLGALLACSPRALILDEPLAGLDVAVSQRLRLLEDLRC
ERGLTVVYVTHDFVGLDEVCPRTVHLRNGALESVSTTAGTS"
6821..7336
/misc_feature
    /gene="ML0848"
    /note="Pfam match to entry PF00005 ABC_tran, ABC
    transporter, score 112.30, E-value 9e-30"
6842..6865
/misc_feature
    /gene="ML0848"
    /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
7109..7153
/misc_feature

Query Match      12.8%; Score 245.2; DB 1; Length 348450;
Best Local Similarity 63.3%; Pred. No. 7.4e-48;
Matches 1134; Conservative 0; Mismatches 652; Indels 6; Gaps 2;

Qy 133  gcgttgattcaaccgcctctccatcatgtattgatattgcacactccacacacactcgct 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39402  CGGTTGGGTTTAAACGGTTGTCCATCATCGACATTCGGCATTCGCATCAGCCGCTGGG 39343

Qy 193  tggggacactgcggatgaaccgcgcgtcaccaatgaactttcaacgggtgagatctacaac 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39342  TGGGACACCGGAGACACCAACCGCTACATCTCTGGTATTCAACGGCGAGATCTACAAC 39283

Qy 253  tacgttgagctgcgtaaagac---tctcggaattgggatatataacctttaactctggc 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39282  TACCTGGAGCTGCGGGACGAGCTGGTCTATCCGGCACGCGCTGTCTTCGCCACCGACGGA 39223

Qy 310  gatggcgagcaaatgttctgcgtttccaccactggggcgagtcgctggtcgagcatctc 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39222  GACGTTGAATCTATCGTCGCGGGCTTCCATTACTGGGGCACCGAGATACTAACAGGCTG 39163

Qy 370  cgcggaattgttcgggaattgcccatttgggatacaaaagggaaaaagtcgttttcttcgcgct 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39162  CGTGCGATGTTCCGCTTTCGCGCTGTGGGACACCATCGCCACGAAATGTTCTGTGCCCGA 39103

Qy 430  gateagttcggcataaagcactgttctaacgaaccaccgagcatggcaacggtttctcc 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39102  GACCCCTTCGGGATCAAAACCACTGTTCTATGCGGACCGAAACGACGACCGCGGTGCC 39043

Qy 490  tcagagaagaagacacattctggagatgcccagagagatgaacttagatctgggcttgat 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39042  ACCGNAAGAAATGCTGCTTGAACCTGGCCCGCAAGTGATCATGATTCGACACCGAGATTGAC 38983

Qy 550  aagcgacacattgagcactacgtggacctgcagtcagtcgcccagagccagatacccttcac 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38982  ACCCGGCACCTACAGCACTACAGGTCCTCGCAGTAGCTGTCGCCGAGCCGAGACGCTGCAT 38923

Qy 610  gcgcagatttcccgcttgagtcaggctgcaccgcaacacgtctgcgggcyg---caag 666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38922  CCGCGGGTGCACCGCACTCGAATCTGGCTGTTACGACACCATATCCCGTCCCGAAACAGCTCGAT 38863

Qy 667  ctggaacagaagcgttacttcaagcctcagttcccagtcacagaaggtcgtaaagggtaag 726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38862  CCAGCGACACCCGCTACTTCTGTCGCCGCAATTTGCGCGCCACACCGATCACCGAGCACT 38803

```

```

Qy 727  gagcaggacctcttcgatcgcatctggcccgagtggttgggaggatagcgtcgaaagcatatg 786
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38802  GAACAGACCCGCTACGGCGAAATCACCCAGTCGTTCGGACTCAGTCGCCAACGACACATG 38743

Qy 787  cgtgccagcgtgacogtaggctcgttcttccttcggggcgcatgactcaacgcgaattgcg 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38742  CGCGCCGACGTCACCGTTGGCTGCTTCTGTCGGCGGATATCGACTCCACAGCCATCGCA 38683

Qy 847  ccgcttcaaaagcgcacaaacccctgacctctcacttcaaccacggttctgcgagcgtgaa 906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38682  GCATGGCGATCCGACACATCCCGGGCTGATTACATTTACCACCGCGTTCGAGCATGAG 38623

Qy 907  ggcactcgagggtcgatgtgtgctgaggagtcgcgcgctgcgattggcgctgagcacatc 966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38622  GGATTTCCGAGATCGACGCTGGCGGCTCTTTCGGCTGAAGCAATCGTTCGCGCCACATC 38563

Qy 967  gtgaagattctcgcctgagaataacgcaacgcggaattcctaagaatcatgttggtacttg 1026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38562  GTGAAGGTGGTTACCCCAACAGAGTTCGTCGCGGCCCTTCGCGGAGATCGTCTGGTATCTC 38503

Qy 1027  gatgacctgtagctgaccccatcattggtccgcgtgtacttcgtgagcggaagacgct 1086
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38502  GACGAGCCAATTCGCGACCCGCGCTGCTTCGCTGCTGTTTTCGTCGCCCGCAACCTGCG 38443

Qy 1087  aagcagctcaaggttgcgtgctgagggcgagcgagatgagctgttcggtggaacacacc 1146
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38442  AAGCAGCTCAAGGTGTTGTCGCGCGAGGCGCGCAGAGCTGTCGCGGCTTTCGCGGCTAC 38383

Qy 1147  atttacaagagcgctatcgctctccattgaagaagatcccttcccactacgtaaa 1206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38382  ATCTATCGAGACCGCTGCTGTGGAAGCCCTTCGACTTACCTACCCGGGCCACTCGGACGC 38323

Qy 1207  ggcctggaaagctcagcaggttctgcagcgcgtcgaagggcgaagtcctctcttgag 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38322  TCGATGGCAAGGTGTCCAATCTCTGCCGAAGTATGTCGGCGCAAGAAATCTGCTACAC 38263

Qy 1267  cgtggctcca tgaocatggaagcgtctactacggaacgctcgtctccttaacttcag 1326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38262  CGCGGATCAGCTCAGCTCGAGGAGCGTACTACGGAATGTCGCGCAATTCGCCGAATTTCT 38203

Qy 1327  cagatgcaacgcttattccatggcgaagcgaatgggaccacgcgaagtaactgca 1386
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38202  CAGTTGCAAGGTGTGCTACCCAGGTTCCACAGGCTTGACACCGGCTGACCCACCGACGTC 38143

Qy 1387  ccgactcactgcacatacccgcaactttgtatccagttagcccccatgcaacacgtgactg 1446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38142  CGGTTGACGCGGAATCGTCGGCTGGGACCCGGTAGCCGCGATGCAGCACATCATGCTG 38083

Qy 1447  ttcacctgga tgcgcggcgacatcctggtcaaggctgacaagatcaacatggcgaaactcc 1506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38082  TTCACCTGGCTACGTGGCGACATTTTGGCCAAGGCGCAAGATGACAAATGGCCAATTTCC 38023

Qy 1507  cttgaagtcgagttccattcttgataagaagatttcaaggttgagagaccattcct 1566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38022  CTGAGACTGCGGGTACCCTTCTTGATCCCAAGTGTTCGCCGCTGGCGTTCGCGTTACCC 37963

Qy 1567  tacgatctgaagattggcacaacggttaccaccagtcacgctgcgcgagggcactgcag 1626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37962  ATGACGGCAAGAGATCACCCGACACCAAGAAATACGACACTGCGGCGCGCATGGAATCC 37903

Qy 1627  attgtccgcctcacgttttgcacgcgaagagctgggcttccctgttcccatgcgcac 1686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37902  ATCGTCCCGGCACACGCTTTCATCGCCCCCAAACTTCGGGTTCCCGGTTCCCAATCCGCAT 37843

Qy 1687  tgggttgcgcgcgagatgagcttctgggttgggcgacagacacattaaaggaatccggtact 1746
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37842  TGGTTCGCGCGCGCGAACTGCTGGAATGGCGCTACGCGCTGGTGAACCTCGTCGAGGCC 37783

Qy 1747  gaagatatcttcaacaagcaggtctgtggtatgttgaaacgagcagccgcgagtgcgtg 1806
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37782  GAACATCTGGTCAACCTAGCCGCGCTGACCCGGATGCTTGACGAGACCCGAGGTGCGATC 37723

Qy 1807  tcagatcaattcccgctgagctgtgagctgttctgttoattatggtgtggaagcaggaatttt 1866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```





```
/translation="MTDVVVVIGIVVAPNGWGAEBYWDATLAGRSLGLPLTRFDSTGY
RAHVAGVNFTEELHPRLPQTDHMTRLALVAAPAEALADAGADPTTMDPDSAGVVT
AASAGFEFGKQELQALSKSPAYSAVQFAFYVNTGQISIRHGRGPSALVAE
QAGGDALAKARRHYREGTALMAGAVDSLCPSWCLLSRSGTVSPRDRPQAFLLPF
DADASGGVPEGGALLVLETPDGAERGHRYRIAGYAATFDRPGSGRPPGLRRA
IDTALADAGIGPSDVVVFADASAEQDRAEAEVLGAVFGPRGPVPTAPKMTGRLL
AGGAPLDVAALLSLRDGVIPPSVHIGRPAHPDLDLVDRPRPARLSALVLARGHG
FNSALVLTs"
2998..3252
/gene="tcsf"
2998..3252
/gene="tcsf"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="BAB12568.1"
/db_xref="GI:9971574"
/translation="MAEFTLDDLIAALKEAAGADEGVAHDDILDVFPFADLGYDSLALF
NTLNTIEREGITLPTDETVEATTPRMLLDVNLGLATAG"
3255..5084
/gene="tcsf"
3255..5084
/gene="tcsf"
/codon_start=1
/transl_table=11
/product="asparagine synthase homolog"
/protein_id="BAB12569.1"
/db_xref="GI:9971575"
/translation="MCGIVGVVDYDAPAEHRDRLLOQMTDTMACRGPDAEGLWFGRAA
GLGHRRLSVIDPEHRGPMTAEHGRGAITFSGEYFNRELRAELTSHGRFRTSCD
TEVLRGLQWALVRLNGALVAFVMDHESEELLVLRDRNGVRPLYYWPTAAGVRF
GSEPKAVLADRTLRSRVDPDGLCEVAVMDVKTPPEAAVFSGLVEVRPQLVVRGSGVRR
QTYWLEAREHTDPLTTIGTVRDLADTVQLVSDVPLGTLSSGLDSSAVTALAA
RSGVRSFSDPRGADAFADAVREADAPVYRELACHVGHADHLEVLDSPLSDPA
VRSVLRATDLPPAYWGMWPSLYLFFRAVRNRYVALSGEAADELFGYRWRNPLA
LRAATFPMLTGPGSARFSGSLFQGLDLKLDLDGYRAARYAEALAEVPLPGESAPE
RMREIHLNLTFRVQTLDRKDRMSMAVLEVRPFCDHRLVEYVFNVPWEMKSFDG
REKSLRAAVADLLPRSVVRVPTPATQDPQEGALRAELAGVLADPDAPVLPDLLD
RERALAAARPAGVSRPYDRGSLSELVLWLSWLAIEYEVTDL"
BASE COUNT      677 a 1873 c 1944 g 694 t
ORIGIN
```

```
Query Match      6.1%; Score 116.6; DB 1; Length 5188;
Best Local Similarity 56.1%; Pred. No. 3.4e-17;
Matches 243; Conservative 0; Mismatches 184; Indels 6; Gaps 1;

81  catgcgcaaccgtggtctgacgatccgggaccttggtgacgacgagcagcgtttgg 140
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  3335 CATGGCCTGCGGGGGCGGACGCCGAGGCGCTCTGGTTCGGCGCGCGGCGTGG 3394

Qy  141 attcaaccgctctccatcattgatttgcacactccaccaccactcggttgggacc 200
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  3395 TCACCGCGCGCTCTCGGTGATCCGCCGAGACGCCGCGCGCGGATGACCGCGGAAC - 3453

Qy  201 tgcggatgaacccgacgcgtacgaatgactttcaacggtgagatctacaactacgttga 260
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  3454 -----ACGAGGGCGCGCGCGCGGATCACCTTCAGCGGGAGATCTACAACCTCCGTGA 3508

Qy  261 gctgcgtaagaagcgtctcggatttgggataatacctttaactcttggcgatggcgagcc 320
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  3509 GCTGCGCGCGGCAACTACCTTCGACGCCGCCACCGGTTTCAGGACGTCCTGCGATACCGAGGT 3568

Qy  321 aattgtgtcggtttccaccactggcgagtcggttcggttcggttcggttcggttcggttc 380
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  3569 CGTCTGCGCGGCTACTCTCAGTGGGGCGCGCGCTGGTTCGAGAGGCTCAACCGGATGTT 3628

Qy  381 cggcattgcccatttgggataaaggaaagtcgcttttcttcttcttcttcttcttcttcttct 440
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  3629 CGCCTTCGCGGCTCTGGGACGAGACACAGGAGAACTCCTCTGCTGCTCGCGACCGGATGGG 3688

Qy  441 catcaagccactgttctacgcaaccaccgagcatggcaccgtgttctctcagagaagaa 500
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```
Db  3689 CGTCAAGCCGCTCTACTACTTGGCGACCGCGCGGCTTTCGGCTTCGGAGCCCAA 3748
Qy  501 gaccatcttgag 513
Db  3749 GGCGCTCCTGGCG 3761

Search completed: September 20, 2002, 07:46:43
Job time: 8395 sec
```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 05:55:13 ; Search time 234.38 Seconds  
(without alignments)  
14064.663 Million cell updates/sec

Title: US-09-786-474-1  
Perfect score: 1920  
Sequence: 1 atgtggcgctcttctggcat.....gctcctaccggtgagcgtt 1920

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

sarched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1          | 1920   | 100.0       | 1920   | 21 | AAA10292    |
| 2          | 1920   | 100.0       | 3825   | 21 | AAA10293    |
| 3          | 1904   | 99.2        | 1920   | 22 | AAH67371    |
| 4          | 1904   | 99.2        | 349980 | 22 | AAH68531    |
| 5          | 1902.4 | 99.1        | 1962   | 22 | AAH71807    |
| c 6        | 116.6  | 6.1         | 30001  | 18 | AAAT61016   |
| c 7        | 116.6  | 6.1         | 30001  | 20 | AAK05110    |
| 8          | 77.4   | 4.0         | 1872   | 22 | AAH77894    |
| c 9        | 74.6   | 3.9         | 349980 | 22 | AAH41225    |

|      |      |     |         |    |           |                    |
|------|------|-----|---------|----|-----------|--------------------|
| 10   | 67.8 | 3.5 | 58857   | 21 | AAA58471  | Nucleotide sequenc |
| 11   | 45.6 | 2.4 | 1470    | 21 | RAC78035  | Human cancer assoc |
| 12   | 45.6 | 2.4 | 1518    | 22 | AA564941  | DNA encoding novel |
| 13   | 45.6 | 2.4 | 3331    | 22 | AA522442  | Human cDNA encodin |
| 14   | 45.6 | 2.4 | 3690    | 22 | AA522678  | Human cDNA encodin |
| 15   | 44.4 | 2.3 | 1836    | 21 | AAAC46946 | Arabidopsis thalia |
| 16   | 42.8 | 2.2 | 2162    | 21 | AAAC50959 | Arabidopsis thalia |
| 17   | 41.6 | 2.2 | 2164    | 21 | AAAC40911 | Arabidopsis thalia |
| 18   | 41.2 | 2.1 | 1664976 | 19 | AAV21209  | Methanococcus jann |
| 19   | 37.8 | 2.0 | 3837    | 23 | ABLO4631  | Drosophila melanog |
| c 20 | 37.6 | 2.0 | 1349    | 21 | AAZ51323  | Human purino recep |
| c 21 | 37.6 | 2.0 | 1349    | 22 | AAH25209  | Nucleotide sequenc |
| c 22 | 37.6 | 2.0 | 1421    | 21 | AAZ51322  | Human purino recep |
| c 23 | 37.6 | 2.0 | 1421    | 22 | AAH25208  | Nucleotide sequenc |
| c 24 | 37.6 | 2.0 | 1436    | 21 | AAZ51321  | Human purino recep |
| c 25 | 37.6 | 2.0 | 1436    | 22 | AAH25207  | Nucleotide sequenc |
| c 26 | 37.6 | 2.0 | 1499    | 21 | AAZ51324  | Human purino recep |
| c 27 | 37.6 | 2.0 | 1499    | 22 | AAH25210  | Nucleotide sequenc |
| c 28 | 36.8 | 1.9 | 1323    | 22 | AAF55439  | Nucleotide sequenc |
| c 29 | 36.8 | 1.9 | 1323    | 23 | AA52741   | E. coli DNA for ce |
| c 30 | 36.8 | 1.9 | 1872    | 23 | AA579318  | DNA encoding novel |
| c 31 | 36.8 | 1.9 | 2047    | 23 | AA593937  | DNA encoding novel |
| c 32 | 36.8 | 1.9 | 3192    | 23 | AA580208  | DNA encoding novel |
| c 33 | 36.8 | 1.9 | 3192    | 23 | AA589446  | DNA encoding novel |
| c 34 | 36.8 | 1.9 | 3192    | 23 | AA593358  | DNA encoding novel |
| c 35 | 36.8 | 1.9 | 19717   | 23 | AA595358  | Propionibacterium  |
| c 36 | 36.6 | 1.9 | 242     | 22 | AAH28183  | Rat differential t |
| c 37 | 36.4 | 1.9 | 1542    | 21 | AAZ60717  | DNA encoding a bet |
| c 38 | 36.4 | 1.9 | 1542    | 22 | AAZ14504  | S. clavuligerus cl |
| c 39 | 36.4 | 1.9 | 2557    | 15 | AAQ62625  | Rat N-acetyl-gluco |
| c 40 | 36.4 | 1.9 | 11604   | 22 | AAZ14501  | Streptomyces clavu |
| c 41 | 36.4 | 1.9 | 15079   | 16 | AAQ91580  | S. clavuligerus cl |
| c 42 | 36.4 | 1.9 | 15079   | 22 | AAZ14499  | Streptomyces clavu |
| c 43 | 36.2 | 1.9 | 1239    | 21 | AAAC5664  | Arabidopsis thalia |
| c 44 | 36.2 | 1.9 | 1242    | 21 | AAAC33977 | Arabidopsis thalia |
| c 45 | 36.2 | 1.9 | 1313    | 22 | AAAF74867 | Leishmania major p |

ALIGNMENTS

RESULT 1  
AAA10292  
ID AAA10292 standard; DNA; 1920 BP.  
XX  
AC AAA10292;  
XX  
03-JUL-2000 (first entry)  
XX  
DE DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.  
XX  
KW Lysozyme insensitivity protein; bacterium; recombinant expression;  
KW amino acid production; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
FH Key Location/Qualifiers  
CDS 1..1920  
FT /\*tag= a  
FT /product= "Corynebacterium glutamicum lysozyme  
insensitivity protein"  
FT /note= "No stop codon given in the specification"  
XX  
PN WO200014241-A1.

16-MAR-2000.

04-SEP-1998; 98WO-JP03981.

04-SEP-1998; 98WO-JP03981.

(KYOW ) KYOWA HAKKO KOGYO KK.

PI Nagai K, Wati M;  
XX WPI: 2000-256989/22.  
DR P-PSDB; AAY87459.  
XX New DNA molecule useful for production of amino acids encodes protein  
PT imparting lysozyme resistance to *Corynebacterium glutamicum* strain -  
XX  
PS Claim 3; Page 33-35; 50pp; Japanese.  
XX This sequence represents the DNA coding sequence encoding a  
CC *Corynebacterium glutamicum* lysozyme insensitive protein. This protein  
CC renders lysozyme-sensitive strains of *Corynebacterium glutamicum*  
CC insensitive to lysozyme. DNA sequences encoding the lysozyme  
CC insensitivity protein can be used to construct a recombinant vector for  
CC the expression of this protein in a host cell. The DNA encoding the  
CC lysozyme insensitive protein can also be mutated, and used to generate  
CC strains of *Corynebacterium glutamicum* in which this protein is  
CC inactivated and which are therefore lysozyme-sensitive. *Corynebacterium*  
CC glutamicum is used for the production of amino acids, in particular  
CC glutamic acid and glutamine.  
Sequence 1920 BP; 425 A; 535 C; 524 G; 436 T; 0 other;

Query Match 100.0%; Score 1920; DB 21; Length 1920;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtgcgccttcttggcatattgactcaaatgggaacgctgaagcattcgttctctca 60  
DB 1 atgtgcgccttcttggcatattgactcaaatgggaacgctgaagcattcgttctctca 60  
QY 61 ctgcagcggccttgcctgcgatgcgcacccgtgctcgtgacgatgcgcgcaacttggcat 120  
DB 61 ctgcagcggccttgcctgcgatgcgcacccgtgctcgtgacgatgcgcgcaacttggcat 120  
QY 121 gacgcgatgacgcttggatttgaacccgctctccatcatgatattgcaactcccaac 180  
DB 121 gacgcgatgacgcttggatttgaacccgctctccatcatgatattgcaactcccaac 180  
QY 181 caaccactgcgttgggaccttgcgatgaacccacgctcagcaatgacttcaacggt 240  
DB 181 caaccactgcgttgggaccttgcgatgaacccacgctcagcaatgacttcaacggt 240  
QY 241 gagatctacaactcgttgcgtgcgttaagagctctcgatttgggatacactttaat 300  
DB 241 gagatctacaactcgttgcgtgcgttaagagctctcgatttgggatacactttaat 300  
QY 301 acttctgcgatggcgagccaattgttctggtttccaccactggcgagtcctggttc 360  
DB 301 acttctgcgatggcgagccaattgttctggtttccaccactggcgagtcctggttc 360  
QY 361 gagcatctcccggaattgttcggcattgccatttgggatacaaaaggaaagtcgttttc 420  
DB 361 gagcatctcccggaattgttcggcattgccatttgggatacaaaaggaaagtcgttttc 420  
QY 421 ctgcgctgatcagctcggcatcaagccactgttctcagcaaccacgagcatggcacc 480  
DB 421 ctgcgctgatcagctcggcatcaagccactgttctcagcaaccacgagcatggcacc 480  
QY 481 gtgttctctcagagaagaacacatttggagatggccggaggagatgaattagatctg 540  
DB 481 gtgttctctcagagaagaacacatttggagatggccggaggagatgaattagatctg 540  
QY 541 ggccttgaagcgcaccattgagcactacgtgaccttgcagtcagtcgtcccgagccagat 600  
DB 541 ggccttgaagcgcaccattgagcactacgtgaccttgcagtcagtcgtcccgagccagat 600  
QY 601 acccttcacgcgagatttccgcttggatgagctgacgtgcaaccagattcgtccgggc 660  
DB 601 acccttcacgcgagatttccgcttggatgagctgacgtgcaaccagattcgtccgggc 660

QY 661 ggcaagctggaacagagcgttacttcaagcctcagttccccagtcacagaaggtcgtaaaag 720  
DB 661 ggcaagctggaacagagcgttacttcaagcctcagttccccagtcacagaaggtcgtaaaag 720  
QY 721 ggttaaggagcaggacctcttcgatacgtattcccaaggtgttggaggatagcgtcgaaaag 780  
DB 721 ggttaaggagcaggacctcttcgatacgtattcccaaggtgttggaggatagcgtcgaaaag 780  
QY 781 catatgcgtgcgcagctgaccgttagcgtctcttcccttcgcgcggtgattgactcaaccgca 840  
DB 781 catatgcgtgcgcagctgaccgttagcgtctcttcccttcgcgcggtgattgactcaaccgca 840  
QY 841 attgcgcgcttgcgaagcgcaaacacctgacctgctcaccttcaccacacgggttctcgag 900  
DB 841 attgcgcgcttgcgaagcgcaaacacctgacctgctcaccttcaccacacgggttctcgag 900  
QY 901 cgtgaagcctactcggaggtgcgattggtcgtcggagtcgcgcgctgcgattggcgctgag 960  
DB 901 cgtgaagcctactcggaggtgcgattggtcgtcggagtcgcgcgctgcgattggcgctgag 960  
QY 961 cacatcgtgaagattgtctcgtcgtgaggaatacgcacacgagattcctcaagatcatgtgg 1020  
DB 961 cacatcgtgaagattgtctcgtcgtgaggaatacgcacacgagattcctcaagatcatgtgg 1020  
QY 1021 tacttggatgacctgtagctgacctatcttgggtcccgctgtacttcgtggcagcgaa 1080  
DB 1021 tacttggatgacctgtagctgacctatcttgggtcccgctgtacttcgtggcagcgaa 1080  
QY 1081 gcaagtaagcagctcaaggttctgctgtcgtcggagggcgagatgagcttgcgttggag 1140  
DB 1081 gcaagtaagcagctcaaggttctgctgtcgtcggagggcgagatgagcttgcgttggag 1140  
QY 1141 tacaccatttcaaaagacgcgctatcgcttctcatttggagaagatcccttccccacta 1200  
DB 1141 tacaccatttcaaaagacgcgctatcgcttctcatttggagaagatcccttccccacta 1200  
QY 1201 cgtaaaagccttggaaagctcagcaaggttctgcagacgcatgaaaggggaagtcctt 1260  
DB 1201 cgtaaaagccttggaaagctcagcaaggttctgcagacgcatgaaaggggaagtcctt 1260  
QY 1261 cttgagcgttgcctcatgacatggaagcgtactactacgcaacgctcgtccctccaat 1320  
DB 1261 cttgagcgttgcctcatgacatggaagcgtactactacgcaacgctcgtccctccaat 1320  
QY 1321 ttgcagcagatgcaacgcgttattccatgggcaaaagcgtgaatgggacacgcggaagtc 1380  
DB 1321 ttgcagcagatgcaacgcgttattccatgggcaaaagcgtgaatgggacacgcggaagtc 1380  
QY 1381 actgcacgactctacgcaacaatccgcacatttgatccagtagcccgcatgcaacacatc 1440  
DB 1381 actgcacgactctacgcaacaatccgcacatttgatccagtagcccgcatgcaacacatc 1440  
QY 1441 gatctgtccactgagtcgcgcgacacatcctggttcaaggctgacaagatcaacatggcg 1500  
DB 1441 gatctgtccactgagtcgcgcgacacatcctggttcaaggctgacaagatcaacatggcg 1500  
QY 1501 aactcccttgagtcgagtttccatttcttgataaggagttttcaaggttgcagagacc 1560  
DB 1501 aactcccttgagtcgagtttccatttcttgataaggagttttcaaggttgcagagacc 1560  
QY 1561 attccttacgactcgaagattgcaacggtaccaccaagtagtcgagtcgagggcactc 1620  
DB 1561 attccttacgactcgaagattgcaacggtaccaccaagtagtcgagtcgagggcactc 1620  
QY 1621 gagcagatttctcgcctcagcttcttcacgcgaagagctgggcttccctgttcccatg 1680  
DB 1621 gagcagatttctcgcctcagcttcttcacgcgaagagctgggcttccctgttcccatg 1680  
QY 1681 cgccactggcttgcgcgcatgagctgttctgggtggcgaggacacaccattaaagaaacc 1740  
DB 1681 cgccactggcttgcgcgcatgagctgttctgggtggcgaggacacaccattaaagaaacc 1740  
QY 1741 ggtactgaagatatcttcaacaagcaggtgtgtgtgatatgtctgaaacgacacgcgcat 1800

Db 1741 ggtactgaagatacttcaacagcagcgctgtctgatatgtctgaacgagcaccgcat 1800  
Qy 1801 ggcgtgcagatcattccctgacgtgacgtgtctgtcattatggtgtgcaaggc 1860  
Db 1801 ggcgtgcagatcattccctgacgtgacgtgtctgtcattatggtgtgcaaggc 1860  
Qy 1861 attttgtgaaacgcgattgacacagatgagagccgctccctaccggtcgagctt 1920  
Db 1861 attttgtgaaacgcgattgacacagatgagagccgctccctaccggtcgagctt 1920

## RESULT 2

AAA10293

ID AAA10293 standard; DNA; 3825 BP.

XX

AC AAA10293;

XX

XX 03-JUL-2000 (first entry)

XX

DE DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.

XX

KW Lysozyme insensitivity protein; bacterium; recombinant expression;

XX

KW amino acid production; ds.

XX

OS Corynebacterium glutamicum.

XX

FH Key Location/Qualifiers

FH 815..2737

FT CDS

FT /\*tag= a

FT /product= "Corynebacterium glutamicum lysozyme

FT insensitivity protein"

XX

XX WO200014241-A1.

XX

XX 16-MAR-2000.

XX

XX 04-SEP-1998; 98WO-JP03981.

XX

XX 04-SEP-1998; 98WO-JP03981.

XX

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX

XX Nagai K, Wati M;

XX

XX WPI; 2000-256989/22.

XX

XX P-PSDB; AAY87459.

XX

XX New DNA molecule useful for production of amino acids encodes protein

XX

XX Imparting lysozyme resistance to Corynebacterium glutamicum strain -

XX

XX Example 2; Page 40-44; 50pp; Japanese.

XX

This sequence represents DNA encoding a Corynebacterium glutamicum  
lysozyme insensitivity protein. This protein renders lysozyme-sensitive  
strains of Corynebacterium glutamicum insensitive to lysozyme. DNA  
sequences encoding the lysozyme insensitivity protein can be used to  
construct a recombinant vector for the expression of this protein in a  
host cell. The DNA encoding the lysozyme insensitivity protein can also  
be mutated, and used to generate strains of Corynebacterium glutamicum in  
which this protein is inactivated and which are therefore lysozyme-  
sensitive. Corynebacterium glutamicum is used for the production of amino  
acids, in particular glutamic acid and glutamine.

XX

XX Sequence 3825 BP; 881 A; 1031 C; 993 G; 920 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 1920; DB 21; Length 3825;

Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 atgtcgcccttcttgcacatgactgcaaatgggaacgctgaagcattcgttccctgca 60

|||||

Db 815 atgtcgcccttcttgcacatgactgcaaatgggaacgctgaagcattcgttccctgca 874  
Qy 61 ctgagcgcccttccatgcacgcgcacgcgtggtcctgacgacgtcgcggcacttgcacat 120  
Db 875 ctgagcgcccttccatgcacgcgcacgcgtggtcctgacgacgtcgcggcacttgcacat 934  
Qy 121 gacgcccagtcagcggttggattcaaccgcctctccatcatttgatattgacacactccac 180  
Db 935 gacgcccagtcagcggttggattcaaccgcctctccatcatttgatattgacacactccac 994  
Qy 181 caaccactgcgttgggacgtcgagatgaaccgcacgcgtcgtacgcaatgacttcaacggt 240  
Db 995 caaccactgcgttgggacgtcgagatgaaccgcacgcgtcgtacgcaatgacttcaacggt 1034  
Qy 241 gacatctacaactacgttggagctcgttaaagagctctcgatttgggatatacactttaat 300  
Db 1055 gacatctacaactacgttggagctcgttaaagagctctcgatttgggatatacactttaat 1114  
Qy 301 acttctggcgatggcgagcccaattgttgcgttttccaccactggggcgagtcctggttc 360  
Db 1115 acttctggcgatggcgagcccaattgttgcgttttccaccactggggcgagtcctggttc 1174  
Qy 361 gacatctcgcggaatgttgcgattgacatttgggatatacacaaggaagtcgttttc 420  
Db 1175 gacatctcgcggaatgttgcgattgacatttgggatatacacaaggaagtcgttttc 1234  
Qy 421 ctgagcggtgacgattcgcgcacacacacacacacacacacacacacacacacacacac 480  
Db 1235 ctgagcggtgacgattcgcgcacacacacacacacacacacacacacacacacacacac 1294  
Qy 481 gtgttctccctcagagaagaagac 540  
Db 1295 gtgttctccctcagagaagaagacacacacacacacacacacacacacacacacacacac 1354  
Qy 541 ggccttgataagcgac 600  
Db 1355 ggccttgataagcgac 1414  
Qy 601 acccttcacgcgagatattcccgcttggagtcagctgacacacacacacacacacacacacac 660  
Db 1415 acccttcacgcgagatattcccgcttggagtcagctgacacacacacacacacacacacacac 1474  
Qy 661 ggcaagctggaacagagcgttacttcaagcctcagttccacagtcacagagtcgtaaaag 720  
Db 1475 ggcaagctggaacagagcgttacttcaagcctcagttccacagtcacagagtcgtaaaag 1534  
Qy 721 ggtaagagcaggacacctcttcgatcgacattgcccaggtgttggaggatagcgtcgaaaaag 780  
Db 1535 ggtaagagcaggacacctcttcgatcgacattgcccaggtgttggaggatagcgtcgaaaaag 1594  
Qy 781 cataatgcgtccgacgtgacgtaggctgttcttcccgcgacattgactcaacgca 840  
Db 1595 cataatgcgtccgacgtgacgtaggctgttcttcccgcgacattgactcaacgca 1654  
Qy 841 attcgcgcgttgcacagcgcac 900  
Db 1655 attcgcgcgttgcacagcgcac 1714  
Qy 901 cgtgaaggctactcggaggtcgatgctgctgagtcgagtcgctgcgctgcgctgcgctgcgag 960  
Db 1715 cgtgaaggctactcggaggtcgatgctgctgagtcgagtcgctgcgctgcgctgcgctgcgag 1774  
Qy 961 cacaatgcgaagattgctgcgctcgaggaatacaccacacacacacacacacacacacacacac 1020  
Db 1775 cacaatgcgaagattgctgcgctcgaggaatacaccacacacacacacacacacacacacacac 1834  
Qy 1021 tacttgatgacctgtagctgacccac 1080  
Db 1835 tacttgatgacctgtagctgacccac 1894  
Qy 1081 gcacgtaagcagctcaagggttgcgtctgctgagggcgagagcgagatgagctgttcggtgga 1140  
Db 1895 gcacgtaagcagctcaagggttgcgtctgctgagggcgagagcgagatgagctgttcggtgga 1954

QY 1141 tacaccattacaagagcgcgtatcgtctgttgcctcatttgagaagatcccttccccacta 1200  
|||||  
Db 1955 tacaccattacaagagcgcgtatcgtctgttgcctcatttgagaagatcccttccccacta 2014  
QY 1201 cgtaaagcctgggaaagctcagaaggtttctgcagacggtcgtggaagggcgaagtccttt 1260  
Db 2015 cgtaaagcctgggaaagctcagaaggtttctgcagacggtcgtggaagggcgaagtccttt 2074  
QY 1261 cttgagcgtggctccatgacacatgaaagagcgtactacggaacgctcgtccctccaat 1320  
Db 2075 cttgagcgtggctccatgacacatgaaagagcgtactacggaacgctcgtccctccaat 2134  
QY 1321 ttcgagcagatcaacgcgtttattccatgggcaaaagcgcgaatgggacacgcgcgaagtc 1380  
Db 2135 ttcgagcagatcaacgcgtttattccatgggcaaaagcgcgaatgggacacgcgcgaagtc 2194  
QY 1381 actgacgcgatacgcacaaatcccgcgaactttgatccagtacgccgcgcatgcaaacactg 1440  
Db 2195 actgacgcgatacgcacaaatcccgcgaactttgatccagtacgccgcgcatgcaaacactg 2254  
1441 gatctgttcaactgatgcgcggcgacatccctgttcaaggctgacaagatcaacatggcg 1500  
2255 gatctgttcaactgatgcgcggcgacatccctgttcaaggctgacaagatcaacatggcg 2314  
QY 1501 aactcccttgagtcgagtgctccattcttggataagggaagttttcaaggttgacagagacc 1560  
Db 2315 aactcccttgagtcgagtgctccattcttggataagggaagttttcaaggttgacagagacc 2374  
QY 1561 attccttacgactgaagatgcgaacggtaccacgaagtaacgctgcgcgaaggaactc 1620  
Db 2375 attccttacgactgaagatgcgaacggtaccacgaagtaacgctgcgcgaaggaactc 2434  
QY 1621 gagcagattgttccgcctcaactgtttgacccgaagagctgggttccctgttcccatg 1680  
Db 2435 gagcagattgttccgcctcaactgtttgacccgaagagctgggttccctgttcccatg 2494  
QY 1681 cgcactggcttgcgcgcgagatgagctgttgcgttggcgccaggacaccattaaaggaatcc 1740  
Db 2495 cgcactggcttgcgcgcgagatgagctgttgcgttggcgccaggacaccattaaaggaatcc 2554  
QY 1741 ggtactgaagatattctcaacagcaggctgtctggtatgtctgatactgaacgagcaccgcgat 1800  
Db 2555 ggtactgaagatattctcaacagcaggctgtctggtatgtctgatactgaacgagcaccgcgat 2614  
QY 1801 ggcgtgtcagatcattcccgctgactgtggactgtctgtcatttattgtgtggcacgac 1860  
Db 2615 ggcgtgtcagatcattcccgctgactgtggactgtctgtcatttattgtgtggcacgac 2674  
QY 1861 atttttggaaaaacgcgattgatccacagattggaggaccgctcctacccggtcgagctt 1920  
Db 2675 atttttggaaaaacgcgattgatccacagattggaggaccgctcctacccggtcgagctt 2734

RESULT 3  
AAH67371  
ID AAH67371 standard; DNA; 1920 BP.  
AC AAH67371;  
XX  
XX  
DT 26-SEP-2001 (first entry)  
XX  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 2406.  
XX  
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX

PF 18-DEC-2000; 2000EP-0127688.  
XX  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI; 2001-376931/40.  
DR P-PSDB; AAG92152.  
XX  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX Claim 8; SEQ ID NO: 2406; 246pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 1920 BP; 422 A; 534 C; 528 G; 436 T; 0 other;

Query Match 99.2%; Score 1904; DB 22; Length 1920;  
Best Local Similarity 99.5%; Pred. NO. 0;  
Matches 1910; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 atgtcgcgccttcttgcatattgactgcaaatgggaacgctgaagcattcctctgca 60  
Db 1 atgtcgcgccttcttgcatattgactgcaaatgggaacgctgaagcattcctctgca 60  
QY 61 ctgagcggccttgcctgcatgcgccacgtggtctcctgacgatgcgcgacttggcat 120  
Db 61 ctgagcggccttgcctgcatgcgccacgtggtctcctgacgatgcgcgacttggcat 120  
QY 121 gacgcgatgcagcgtttggattcaaccgcctctccatcattgatattgcacactccac 180  
Db 121 gacgcgatgcagcgtttggattcaaccgcctctccatcattgatattgcacactccac 180  
QY 181 caaccactcgttggggacctgcggatgaacccgaccgctacgcaatgacttcaacggt 240  
Db 181 caaccactcgttggggacctgcggatgaacccgaccgctacgcaatgacttcaacggt 240  
QY 241 gagatctacaactacgttgagctgcgttaagagcctcggatttgggatatatcctttaa 300  
Db 241 gagatctacaactacgttgagctgcgttaagagcctcggatttgggatatatcctttaa 300  
QY 301 acttctggcgatggcgagcccaattgttgcgttttcccaccactggggcgagtcctggtc 360  
Db 301 acttctggcgatggcgagcccaattgttgcgttttcccaccactggggcgagtcctggtc 360  
QY 361 gagcatctccgcggaatttccggcattgccaatttgggatacaaaagaaagtcgttttc 420  
Db 361 gagcatctccgcggaatttccggcattgccaatttgggatacaaaagaaagtcgttttc 420  
QY 421 cttgcgctgatcagttcggcatcaaacgactgttctcagcaaacaccagcagatggcacc 480  
Db 421 cttgcgctgatcagttcggcatcaaacgactgttctcagcaaacaccagcagatggcacc 480

```
QY 481 gtgttctctcagagaagaccattcttggagatggtccgagagagatgaatctagatctg 540
Db 481 gtgttctctcagagaagaccattcttggagatggtccgagagagatgaatctagatctg 540
QY 541 ggccttgataagcgaccattgagcactacgttgagcctgcagtacgtgcccgagccagat 600
Db 541 ggccttgataagcgaccattgagcactacgttgagcctgcagtacgtgcccgagccagat 600
QY 601 acccttcacgcgagatttcccgcttggagtcagggtgcacgcacaaagttctccgggc 660
Db 601 acccttcacgcgagatttcccgcttggagtcagggtgcacgcacaaagttctccgggc 660
QY 661 ggaagctggaacagaagcgttacttcaagcctcagttcccaagcagaaggtcgttaaag 720
Db 661 ggaagctggaacagaagcgttacttcaagcctcagttcccaagcagaaggtcgttaaag 720
QY 721 ggtaaagagcaggaacctcttcgatcgattgccaggtgttgagagatagcgtcgaaaag 780
Db 721 ggtaaagagcaggaacctcttcgatcgattgccaggtgttgagagatagcgtcgaaaag 780
QY 781 catatcgctgcagcgtgacgttaggctcgattgcttcttcggcgccgattgactcaacgcga 840
Db 781 catatcgctgcagcgtgacgttaggctcgattgcttcttcggcgccgattgactcaacgcga 840
QY 841 attgcgccccttgcgaagcgccacaacacctgacctcacttcacccacgggtttcgag 900
Db 841 attgcgccccttgcgaagcgccacaacacctgacctcacttcacccacgggtttcgag 900
QY 901 cgtgaaggtctactcggaggtcgatgtggctgcggagtcggcgctgcgattggcgctgag 960
Db 901 cgtgaaggtctactcggaggtcgatgtggctgcggagtcggcgctgcgattggcgctgag 960
QY 961 cacatcgtgaagattgtctgcgtcgtggaatacgcgaacgcgatttcttaagatcatgtgg 1020
Db 961 cacatcgtgaagattgtctgcgtcgtggaatacgcgaacgcgatttcttaagatcatgtgg 1020
QY 1021 tacttgatgatcctgtagctgaccatcatgtgtcccgctgtacttctggtgcagcgagaa 1080
Db 1021 tacttgatgatcctgtagctgaccatcatgtgtcccgctgtacttctggtgcagcgagaa 1080
QY 1081 gcacgtaagcagctcaaggttgctgtctgcggagggcgagatgagctgttcggtgga 1140
Db 1081 gcacgtaagcagctcaaggttgctgtctgcggagggcgagatgagctgttcggtgga 1140
QY 1141 tacacatttacaagaagcgcgtatcgctgtccatttgagaagatcccttcccacta 1200
Db 1141 tacacatttacaagaagcgcgtatcgctgtccatttgagaagatcccttcccacta 1200
QY 1201 cgtaaagggctgggaagcctcagcaaggttctgcgagggcgagatgaagggcaagtccctt 1260
Db 1201 cgtaaagggctgggaagcctcagcaaggttctgcgagggcgagatgaagggcaagtccctt 1260
QY 1261 ctgtgagctggctccatgacctggaagcgcgtactacggcaacgcctgcctcttcaat 1320
Db 1261 ctgtgagctggctccatgacctggaagcgcgtactacggcaacgcctgcctcttcaat 1320
QY 1321 ttcgagcagatgcaacgcgttatccatgggcaaaagcggaaatgggacacaccccgaaagtc 1380
Db 1321 ttcgagcagatgcaacgcgttatccatgggcaaaagcggaaatgggacacaccccgaaagtc 1380
QY 1381 actgcaccgatctacgcacaatcccgaactttgatccagtagcccgcatgcaaacctg 1440
Db 1381 actgcaccgatctacgcacaatcccgaactttgatccagtagcccgcatgcaaacctg 1440
QY 1441 gatctgtacacctggatgcgcgcgacatcctggtcaagctgacaagatcaacatggcg 1500
Db 1441 gatctgtacacctggatgcgcgcgacatcctggtcaagctgacaagatcaacatggcg 1500
QY 1501 aactccttgagctgcgagttccattcttgataaaggaagtttcaaggttgagagacc 1560
Db 1501 aactccttgagctgcgagttccattcttgataaaggaagtttcaaggttgagagacc 1560
```

## RESULT 4

AAH68531  
ID AAH68531 standard; DNA; 349980 BP.

XX AC AAH68531;

XX AC (first entry)

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 7066.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

XX mutation point of a gene, measuring expression of a gene, analysing

XX expression profile or pattern of a gene and identifying homologous gene

XX Disclosure; SEQ ID NO: 7066; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
are useful for identifying the mutation point of a gene derived from a  
mutant of coryneform bacterium, measuring expression amount and  
analysing the expression profile or expression pattern of a gene derived  
from coryneform bacterium, and identifying a homologue of a gene derived  
from coryneform bacterium. Coryneform bacteria are useful for producing  
amino acids, nucleic acids, vitamins, saccharides and organic acids,  
particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;

Query Match 99.2%; Score 1904; DB 22; Length 349980;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1910; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 atgtcgcccttcttgccatattgactgcaaatgggaacgctgaagcattgcttcctaca 60  
|||||  
Db 228516 atgtcgcccttcttgccatattgactgcaaatgggaacgctgaagcattgcttcgca 228575  
|||||

Qy 61 ctcgagcgcccttcccatgcatgcgcacacgtgtctctgacgatgccggaacttggcat 120  
|||||  
Db 228576 ctcgagcgcccttcccatgcatgcgcacacgtgtctctgacgatgccggaacttggcat 228635  
|||||

Qy 121 gacgcgatgacgcttggattcaaccgctctccatcattgattgacacatcccac 180  
|||||  
228636 gacgcgatgacgcttggattcaaccgctctccatcattgattgacacatcccac 228695  
|||||

Qy 181 caaccactcgcttggggacctcgccgatgaaccgacgctacgcaaatgacttcaacggt 240  
|||||  
Db 228696 caaccactcgcttggggacctcgccgatgaaccgacgctacgcaaatgacttcaacggt 228755  
|||||

Qy 241 gagatctcaactacgttgcgtgcgtaagagctctcgatttgggatacactttaat 300  
|||||  
Db 228756 gagatctcaactacgttgcgtgcgtaagagctctcgatttgggatacactttaat 228815  
|||||

Qy 301 acttctgcgatggcgagcccaattgtctggtttccaccactggggcgagtcctgtgtc 360  
|||||  
Db 228816 acttctgcgatggcgagcccaattgtctggtttccaccactggggcgagtcctgtgtc 228875  
|||||

Qy 361 gagcatctccgcgggaattgtcggcaattgccatttgggatacaaaagaaagtcgttttc 420  
|||||  
Db 228876 gagcatctccgcgggaattgtcggcaattgccatttgggatacaaaagaaagtcgttttc 228935  
|||||

Qy 421 ctgcgcgtgatcagcttcggcatacgaacactgttctacgcaaccacgacgatgcacc 480  
|||||  
Db 228936 ctgcgcgtgatcagcttcggcatacgaacactgttctacgcaaccacgacgatgcacc 228995  
|||||

Qy 481 gtgtctctctcagaaagaagaccattcttgagatggcgaggagatgaattagatctg 540  
|||||  
Db 228996 gtgtctctcagaaagaagaccattcttgagatggcgaggagatgaattagatctg 229055  
|||||

Qy 541 ggccttgataagccacattgagcactacgttggacctggacctgcccagagccagat 600  
|||||  
b 229056 ggccttgataagccacattgagcactacgttggacctggacctgcccagagccagat 229115  
|||||

Qy 601 acccttcacgcgcagatttcccgcttgagtcagctgcacgcacacagttcgtccggc 660  
|||||  
Db 229116 acccttcacgcgcagatttcccgcttgagtcagctgcacgcacacagttcgtccggc 229175  
|||||

Qy 661 ggcaagctggaaacagaagcttacttcaagctcagttcccgatcacagaagtgctgaag 720  
|||||  
Db 229176 ggcaagctggaaacagaagcttacttcaagctcagttcccgatcacagaagtgctgaag 229235  
|||||

Qy 721 ggttaagagacgaacctcttcgatcgacattgcccaagtgcttggaggaagcgcgaagaag 780  
|||||  
Db 229236 ggttaagagacgaacctcttcgatcgacattgcccaagtgcttggaggaagcgcgcgaagaag 229295  
|||||

Qy 781 catatgctgcccagctgacacgttagctcttcccttccgcggcattgactcaaccgca 840  
|||||  
Db 229296 catatgctgcccagctgacacgttagctcttcccttccgcggcattgactcaaccgca 229355  
|||||

Qy 841 attgcgcgcttggaaaagcccaaacacctgactgctcacttcaaccacccggtttcgag 900  
|||||  
Db 229356 attgcgcgcttggaaaagcccaaacacctgactgctcacttcaaccacccggtttcgag 229415  
|||||

Qy 901 cgtgaaggctactcggaggttcgatgtggctcggagctccgcgctgcgattggcgctgag 960  
|||||

RESULT 5  
AAF71807  
ID AAF71807 standard; DNA; 1962 BP.  
XX  
AC AAF71807;

Db 229416 cgtgaaggctactcgtgaggttcgatgtggtcggagtcgcgcgtcgcgtatggcgctgag 229475  
|||||

Qy 961 cacatcgtgaagattgtctcgcctgaggaataacgccaacgggattcctaagaatcatgtgg 1020  
|||||

Db 229476 cacatcgtgaagattgtctcgcctgaggaataacgccaacgggattcctaagaatcatgtgg 229535  
|||||

Qy 1021 tacttggatgacctgtagctgaccatcaatttggctccgcgtgtacttctgtggcagcgaaa 1080  
|||||

Db 229536 tacttggatgacctgtagctgaccatcaatttggctccgcgtgtacttctgtggcagcgaaa 229595  
|||||

Qy 1081 gcaagtgaagcagctcaaggttctgtctgtcggagggcgagatgagctgttcgttga 1140  
|||||

Db 229596 gcaagtgaagcagctcaaggttctgtctgtcggagggcgagatgagctgttcgttga 229655  
|||||

Qy 1141 tacaccatttcaaaagagcgcctatcgctgtctccatttggagaagatcccttcccacta 1200  
|||||

Db 229656 tacaccatttcaaaagagcgcctatcgctgtctccatttggagaagatcccttcccacta 229715  
|||||

Qy 1201 cgtaaagccttgggaaagctcagcaaggttcttgcagacggtcgaatgaaggcaagtcctt 1260  
|||||

Db 229716 cgtaaagccttgggaaagctcagcaaggttcttgcagacggtcgaatgaaggcaagtcctt 229775  
|||||

Qy 1261 cttgagcgtggctccatgacatggaagagcgtactacggaacgctcgcctcctcaat 1320  
|||||

Db 229776 cttgagcgtggctccatgacatggaagagcgtactacggaacgctcgcctcctcaat 229835  
|||||

Qy 1321 ttgcagcagatgcacgcgttatttcattcgtggcaagcgtgaatgggacacgcggaagtc 1380  
|||||

Db 229836 ttgcagcagatgcacgcgttatttcattcgtggcaagcgtgaatgggacacgcggaagtc 229895  
|||||

Qy 1381 actgcaccgatactacgcacaatcccgcaacttgcagtagccccgcgatgcacaacctg 1440  
|||||

Db 229896 actgcaccgatactacgcacagtcctccgcaacttgcagtagccccgcgatgcacaacctg 229955  
|||||

Qy 1441 gatctgttcaactggatgcgcggcgacatcctcgttgcaggctgacaagaatacaatggcg 1500  
|||||

Db 229956 gatctgttcaactggatgcgcggcgacatcctcgttgcaggctgacaagaatacaatggcg 230015  
|||||

Qy 1501 aactcccttgagctgcgagttccattcttgcgtaaggaagttttcaaggttgcagagacc 1560  
|||||

Db 230016 aactcccttgagctgcgagttccattcttgcgtaaggaagttttcaaggttgcagagacc 230075  
|||||

Qy 1561 attccttacgactcgaagattgcgaacggtaccaccaagtaogcgtgcgagggcactc 1620  
|||||

Db 230076 attccttacgactcgaagattgcgaacggtaccaccaagtaogcgtgcgagggcactc 230135  
|||||

Qy 1621 gagcagattgttccgcctcagcttttgcacgcgaagaagctgggcttccctgttcccaatg 1680  
|||||

Db 230136 gagcagattgttccgcctcagcttttgcacgcgaagaagctgggcttccctgttcccaatg 230195  
|||||

Qy 1681 cgccactggcttgcgcgcgagtgagctgttcggttggcgcgaggaacaccattaaagaaatcc 1740  
|||||

Db 230196 cgccactggcttgcgcgcgagtgagctgttcggttggcgcgaggaacaccattaaagaaatcc 230255  
|||||

Qy 1741 ggtactgaagatatcttcaacaagcaggcgtgtgctggtatgctgaacgagcaccgcgat 1800  
|||||

Db 230256 ggtactgaagatatcttcaacaagcaggcgtgtgctggtatgctgaacgagcaccgcgat 230315  
|||||

Qy 1801 ggcgtgtcagatcattcccgctcagctgtggactgttctgtcatttattggtggcagcgc 1860  
|||||

Db 230316 ggcgtgtcagatcattcccgctcagctgtggactgttctgtcatttattggtggcagcgc 230375  
|||||

Qy 1861 atttttggaaaaaccgcatgtaccacagattgaggaccgctcctaccggtcagctt 1920  
|||||

Db 230376 atttttggaaaaaccgcatgtaccacagattgaggaccgctcctaccggtcagctt 230435  
|||||

XX 30-APR-2001 (first entry)  
DT Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:109.  
XX  
DE  
XX  
XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
XX WO200100843-A2.  
PN  
XX  
XX  
PD 04-JAN-2001.  
XX  
XX  
XX 23-JUN-2000; 2000WO-IB00923.  
XX  
XX 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041394.  
PR 03-SEP-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.

PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
XX (BADI ) BASF AG.  
PA  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;  
PI  
XX  
XX WPI; 2001-137957/14.  
DR P-PSDB; AAB79688.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX  
XX Claim 3; Page 324-327; 1737pp; English.  
XX  
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC MP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polyketides and enzymes.  
XX  
XX Sequence 1962 BP; 435 A; 544 C; 534 G; 449 T; 0 other;  
SQ  
  
Query Match 99.1%; Score 1902.4; DB 22; Length 1962;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1909; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
  
QY 1 atgtcgccgtcttcttgacatattgactgcaaatgggaacgtgaagcattcggtcttcgca 60  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
20 atgtcgccgtcttcttgacatattgactgcaaatgggaacgtgaagcattcggtcttcgca 79  
QY 61 ctcgagcgccgttgcacatgcgcacccgtggtcctgacatgcgcacattgacat 120  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
80 ctcgagcgccgttgcacatgcgcacccgtggtcctgacatgcgcacattgacat 139  
QY 121 gacgcgcatgcagcttggattcaaccgcctctccatcattgattgacacattccac 180  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
140 gacgcgcatgcagcttggattcaaccgcctctccatcattgattgacacattccac 199  
QY 181 caaccactgcgttgggacatgcgcacccgcacccgcacccgcacccgcacccggt 240  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
200 caaccactgcgttgggacatgcgcacccgcacccgcacccgcacccgcacccggt 259  
QY 241 gagatctacaactacgttgcagtcgctaaagagctctcggatttggatatacatttcaat 300  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
260 gagatctacaactacgttgcagtcgctaaagagctctcggatttggatatacatttcaat 319  
QY 301 actcttgccgatggcgagcccaattgttgcgtttccaccactgggagcgtccgtggtc 360  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
320 actcttgccgatggcgagcccaattgttgcgtttccaccactgggagcgtccgtggtc 379  
QY 361 gagcatctccgcggaatggttcggcattgccatttgggatacaaaaggaagtcgtcttcc 420  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
380 gagcatctccgcggaatggttcggcattgccatttgggatacaaaaggaagtcgtcttcc 439  
QY 421 ctgtcgcggtgatcagttcggcatcaagccactgttctacgaaccaccagcagtcgacc 480  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
440 ctgtcgcggtgatcagttcggcatcaagccactgttctacgaaccaccagcagtcgacc 499  
QY 481 gtgttctctcagagaagaccatttggagatggcccgagagatgaattcagatctg 540  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
500 gtgttctctcagagaagaccatttggagatggcccgagagatgaattcagatctg 559  
QY 541 ggccttgataagcgcaccatttgagcactacgttgacactgcagtcagtcgcccagcagat 600  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
560 ggccttgataagcgcaccatttgagcactacgttgagacttgagactgcgcccagcagat 619





Db 22834 CATGCGCTGCGGGGCGGACGCGGAGGCGCTCTGTTGCGCGCGCGCGGCTCGG 22775  
QY 141 attcaacggcctctccatcattgatattgcacactcccacaaacactcggtgggacc 200  
Db 22774 TCACCGCGCGCTGCTGGTATCGACCCCGAGCACGCGCGGACGCGGATACCGCGGAAC- 22714  
QY 201 tgcggatgaacccgacgctacgaatgactttcaacggtgagatctacaactacgttga 260  
Db 22715 -----ACGAGGGCGCGGCGCGGATCACCTTCACGCGGAGATCTACAACCTCCGTGA 22661  
QY 261 gctgcgtaaaagactcctggattggatatacactttaatactcttgcgctgacgttcgg 320  
Db 22660 GCTGCGCGCGGACCTACCTCGCACGCGGACCGGTTCTGCTGCTCGCGGACCGGATGG 22481  
QY 321 aattgtgtcgtttccacactggcgagtcggtggtcgagcatctccgcggaatgtt 380  
Db 22600 CGTCTGCGCGGCTACTACTGCGCGGACCGCGCGGCTGCGGCTCGGAGGCCAA 22421  
QY 501 gaccatctggag 513  
Db 22420 GGCGGCTCTGGCG 22408

RESULT 7  
ID AAX05110/C  
XX AAX05110 standard; DNA; 30001 BP.  
AC AAX05110;  
XX  
DT 07-APR-1999 (first entry)  
XX  
DE S. aureofaciens DNA in cosmid clones LP2-127 and LP2-128.  
XX  
KW Chlortetracycline; tetracycline; biosynthetic; actinomycete;  
KW S. arisefuscus; S. ambofaciens; antibiotic resistance gene;  
KW S. lividans; S. aureofaciens; cosmid clone; antibiotic; ss.  
XX  
OS Streptomyces aureofaciens.  
XX  
X US5866410-A.  
XX  
PD 02-FEB-1999.  
XX  
PF 07-JUN-1995; 95US-0474933.  
XX  
PR 22-SEP-1993; 93US-0125468.  
PR 26-JUL-1990; 90US-0558039.  
PR 26-JUL-1990; 90US-0558040.  
PR 15-JAN-1992; 92US-0821109.  
PR 15-JAN-1992; 92US-0821419.  
PR 07-JUN-1995; 95US-0474933.  
XX  
PA (AMCY ) AMERICAN CYANAMID CO.  
XX  
PI Fantini SE, Lotvin JA, Ryan MJ, Strathy N;  
XX  
XX WPI; 1999-141936/12.  
XX  
XX Two plasmids for cloning the biosynthetic pathways of tetracycline,  
PT chlortetracycline, and their derivatives - comprise an origin of  
PT replication, cohesive end sites and optionally, an antibiotic  
PT resistance gene  
XX  
PS Disclosure; Fig 4A-L; 39pp; English.  
XX

CC The invention relates to combination of two plasmids for cloning the  
CC genes encoding the biosynthetic production pathways of chlortetracycline,  
CC tetracycline, or their derivatives. One plasmid contains an origin of  
CC replication, an actinomycete (such as Streptomyces lividans,  
CC S. arisefuscus, or S. ambofaciens) active antibiotic resistance gene and  
CC three or more tandem cohesive end sites. The second plasmid contains an  
CC actinomycete active origin of replication and three or more tandem  
CC cohesive end sites. The invention can be used to clone a large amount of  
CC genetic material for the heterologous production of the antibiotics,  
CC chlortetracycline, tetracycline and their analogues. A cluster of genes  
CC can be inserted into the plasmid allowing a biosynthetic pathway to be  
CC transferred in its entirety to a heterologous host. The invention  
CC specifically relates to cloning of the entire tetracycline and  
CC chlortetracycline from Streptomyces aureofaciens and its expression in a  
CC heterologous host such as S. lividans. The present sequence represents a  
CC S. aureofaciens DNA sequence from the cosmid clones designated LP2-127  
CC and LP2-128.  
XX  
SQ Sequence 30001 BP; 4055 A; 10939 C; 11019 G; 3988 T; 0 other;  
Query Match 6.1%; Score 116.6; DB 20; Length 30001;  
Best Local Similarity 56.1%; Pred. No. 2.6e-23;  
Matches 243; Conservative 0; Mismatches 184; Indels 6; Gaps 1;  
QY 81 catgcgccaccgtgtcctgacgatgccggcacttggcatgacgcccgttgg 140  
Db 22834 CATGGCTGCGGGGCGGACGCGGAGGCGCTCTGTTGCGCGCGCGGCTCGG 22775  
QY 141 attcaacggcctctccatcattgatattgcacactcccacaaacactcggtgggacc 200  
Db 22774 TCACCGCGCGCTGCTGGTATCGACCCCGAGCACGCGCGGACCGGATGACCGCGGAAC- 22714  
QY 201 tgcggatgaacccgacgctacgaatgactttcaacggtgagatctacaactacgttga 260  
Db 22715 -----ACGAGGGCGCGGCGCGGATCACCTTCAGCGGAGATCTACAACCTCCGTGA 22661  
QY 261 gctgcgtaaaagactcctggattggatatacactttaatactcttgcgctgacgttcgg 320  
Db 22660 GCTGCGCGCGGACCTACCTCGCACGCGGACCGGTTCTGCTGCTCGCGGACCGGATGG 22481  
QY 321 aattgtgtcgtttccacactggcgagtcggtggtcgagcatctccgcggaatgtt 380  
Db 22600 CGTCTGCGCGGCTACTACTGCGCGGACCGCGCGGCTGCGGCTCGGAGGCCAA 22421  
QY 501 gaccatctggag 513  
Db 22420 GGCGGCTCTGGCG 22408  
RESULT 8  
ID AAF77894  
XX AAF77894 standard; DNA; 1872 BP.  
XX  
AC AAF77894;  
XX  
DT 04-JUN-2001 (first entry)  
XX  
DE Quorum sensing controlled gene qscI37 ORF.  
XX  
KW Quorum sensing; antibacterial; bacterial signalling;  
KW opportunistic pathogen; immunocompromised; burn; cystic fibrosis;  
KW immunosuppressive therapy; AIDS; ss.  
XX  
OS Pseudomonas aeruginosa.







|          |  |                          |
|----------|--|--------------------------|
| RESULT   | 11   |                          |
| AAC78035 |  |                          |
| ID       | AAC78035   | standard; cDNA; 1470 BP. |
| XX       |  |                          |
| AC       | AAC78035;  |                          |
| XX       |  |                          |
| DT       | 08-FEB-2001  | (first entry)            |
| XX       |  |                          |
| DE       | Human cancer associated gene sequence  | SEQ ID NO:429.           |
| XX       |  |                          |
| KW       | Human; cancer associated gene; cancer antigen; detection; cancer;                |                          |
| KW       | diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;               |                          |
| KW       | antidiabetic; antisthmatic; antirheumatic; antithrombotic; antiviral;            |                          |
| KW       | antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;             |                          |
| KW       | dermatological; neuroprotective; thrombolytic; coagulant; nontropic;             |                          |
| KW       | vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;           |                          |
| KW       | immune disorder; haematopoietic cell disorder; autoimmune disorder;              |                          |
| KW       | allergic reaction; graft versus host disease; organ rejection;                   |                          |
| KW       | haemostatic; thrombolytic; cardiovascular disorder; infection;                   |                          |
| KW       | neurological disease; drug screening; ss.  |                          |
| OS       | Homo sapiens.  |                          |
| XX       |  |                          |
| PN       | W0200055350-A1.  |                          |
| XX       |  |                          |
| PD       | 21-SEP-2000.   |                          |
| XX       |  |                          |
| PF       | 08-MAR-2000; 200WO-US05882.  |                          |
| XX       |  |                          |
| PR       | 12-MAR-1999; 99US-0124270.   |                          |
| XX       |  |                          |
| PA       | (HUMA-) HUMAN GENOME SCI INC.  |                          |
| XX       |  |                          |
| PI       | Rosen CA, Ruben SM;  |                          |
| XX       |  |                          |
| DR       | WPI; 2000-587533/55.   |                          |
| XX       |  |                          |
| DR       | P-PSDB; AAB43826.  |                          |
| XX       |  |                          |
| PT       | Novel isolated nucleic acids comprising sequences encoding peptides              |                          |
| XX       |  |                          |
| PS       | useful for treating or diagnosing e.g. cancer -                                  |                          |
| XX       |  |                          |
| CC       | Claim 1; Page 970; 2352pp; English.  |                          |
| CC       | AAC77607 to AAC78448 encode the human cancer associated proteins given           |                          |
| CC       | in AAB43398 to AAB44239. The proteins can have activities based on the           |                          |
| CC       | tissues and cells the genes are expressed in. Example of activities              |                          |
| CC       | include: cytostatic; proliferative; vulnerable; immunomodulator;                 |                          |
| CC       | antidiabetic; antisthmatic; antirheumatic; antithyroid;                          |                          |
| CC       | antiinflammatory; antichyroid; antiallergic; antibacterial; antiviral;           |                          |
| CC       | dermatological; neuroprotective; cardiac; thrombolytic; coagulant;               |                          |
| CC       | nontropic; vasotropic; antipsoriatic and antiangiogenic. The                     |                          |
| CC       | polynucleotides and polypeptides can be used for preventing, treating or         |                          |
| CC       | ameliorating medical conditions and diagnosing pathological conditions           |                          |
| CC       | the present invention may be used to treat immune disorders and antagonists from |                          |
| CC       | or inhibiting the proliferation, differentiation or mobilisation of              |                          |
| CC       | immune cells, to treat disorders of haematopoietic cells, autoimmune             |                          |
| CC       | disorders, allergic reactions, graft versus host disease and organ               |                          |
| CC       | rejection, modulate haemostatic or thrombolytic activity, modulate               |                          |
| CC       | inflammation, cancers, cardiovascular disorders, neurological disease and        |                          |
| CC       | bacterial or viral infections. The peptides, nucleotides, antibodies, and        |                          |
| CC       | agonists and antagonists may be also be used in drug screens. AAC78449 to        |                          |
| CC       | AAC78457 and AAB44240 represent sequences used in the exemplification of         |                          |
| CC       | the present invention.   |                          |
| XX       |  |                          |
| SQ       | Sequence 1470 BP; 336 A; 380 C; 470 G; 277 T; 7 other;                           |                          |

|                       |                 |                  |           |              |
|-----------------------|-----------------|------------------|-----------|--------------|
| Query Match           | 2.48;           | Score 45.6;      | DB 21;    | Length 1470; |
| Best Local Similarity | 49.28;          | Pred. No. 0.007; |           |              |
| Matches 120;          | Conservative 0; | Mismatches 124;  | Indels 0; | Gaps         |

|               |   |  |      |
|---------------|---|--|------|
| Qy            | 1409  | actttgatccagtagccgcgcatgcacaacctggatctgttctcaacttgatgcgcgcgacga  | 1466 |
| Db            | 451   |  | 510  |
| Qy            | 1469  | tcttggtccaaggctgacaagatacaacatgcggaactccttgagctgcgagttccattct    | 1528 |
| Db            | 511   |  | 570  |
| Qy            | 1529  | tggataaggaaagtcttcagggttcgacagagaccattccttacgatactgaagattgccaacg | 1588 |
| Db            | 571   |  | 630  |
| Qy            | 1589  | gtaccaccaagaatcacgcgtcgcagggaactcgacagatggttcgcctcacagttttgc     | 1648 |
| Db            | 631   |  | 690  |
| Qy            | 1649  | acgcg 1652   |      |
| Db            | 691   | agcg 694   |      |
| <br>RESULT 12 |   |  |      |
| AAS64941      |   |  |      |
| ID            | AAS64941  | standard; cDNA; 1518 BP.   |      |
| XX            | AC  | AAS64941;  |      |
| XX            | DT  | 13-FEB-2002 (first entry)  |      |
| XX            | DE  | DNA encoding novel human diagnostic protein #745.                |      |
| KW            | Human; chromosome mapping; gene mapping; gene therapy; forensic;        |  |      |
| KW            | food supplement; medical imaging; diagnostic; genetic disorder; ss.     |  |      |
| OS            | Homo sapiens.   |  |      |
| XX            | PN  | W0200175067-A2.  |      |
| XX            | PD  | 11-OCT-2001.   |      |
| XX            | PF  | 30-MAR-2001; 2001WO-US08631.                                     |      |
| XX            | PR  | 31-MAR-2000; 2000US-0540217.                                     |      |
| PR            | 23-AUG-2000; 2000US-0649167.  |  |      |
| XX            | PA  | (HYSE-) HYSEQ INC.   |      |
| XX            | PI  | Drmanac RT, Liu C, Tang YT;                                      |      |
| XX            | DR  | WFI; 2001-639362/73.   |      |
| DR            | P-PSDB; ABG00754.   |  |      |
| PT            | New isolated polynucleotide and encoded polypeptides, useful in         |  |      |
| PT            | diagnostics, forensics, gene mapping, identification of mutations       |  |      |
| PT            | responsible for genetic disorders or other traits and to assess         |  |      |
| PT            | biodiversity -  |  |      |
| XX            | Claim 1; SEQ ID NO 745; 103pp; English.                                 |  |      |
| XX            | The invention relates to isolated polynucleotide (I) and                |  |      |
| CC            | polypeptide (II) sequences. (I) is useful as hybridisation probes,      |  |      |
| CC            | polymerase chain reaction (PCR) primers, oligomers, and for chromosome  |  |      |
| CC            | and gene mapping, and in recombinant production of (II). The            |  |      |
| CC            | polynucleotides are also used in diagnostics as expressed sequence tags |  |      |
| CC            | for identifying expressed genes. (I) is useful in gene therapy techniq  |  |      |
| CC            | to restore normal activity of (II) or to treat disease states involv    |  |      |
| CC            | (II). (II) is useful for generating antibodies against it, detecting o  |  |      |
| CC            | quantitating a polypeptide in tissue, as molecular weight markers and   |  |      |
| CC            | a food supplement. (II) and its binding partners are useful in medic    |  |      |
| CC            | imaging of sites expressing (II). (I) and (II) are useful for treating  |  |      |
| CC            | disorders involving aberrant protein expression or biological activ     |  |      |
| CC            | The polypeptide and polynucleotide sequences have applications in       |  |      |

```
Query Match      2.4%; Score 45.6; DB 21; Length 1470;
Best Local Similarity 49.2%; Pred. NO. 0.007;
Matches 120; Conservative 0; Mismatches 124; Indels 0; Gaps 0
```



KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.  
 XX Homo sapiens.

OS

XX WO200155437-A2.

PN

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

PI WPI; 2001-451939/48.

DR P-PSDB; AAU14373.

XX

PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage -  
 XX

PS Claim 1; Page 673-674; 894pp; English.

CC The invention relates to polynucleotides encoding novel human  
 CC proteins or their active domains. The polypeptides, polynucleotides and  
 CC antibodies raised against the polypeptides are used in a method of  
 CC treatment of a mammal and prevention of disorders caused by the aberrant  
 CC protein expression or activity. The polypeptides can be used as  
 CC molecular weight markers, food supplements, and in antibody production.  
 CC The polypeptides are used to identify compounds which bind to the  
 CC polypeptides. Polynucleotides of the invention are used as probes and  
 CC primers, for sequencing, for chromosome or gene mapping, in the  
 CC production of recombinant proteins, and in generating anti-sense DNA or  
 CC RNA and in gene therapy. Polypeptides of the invention can be used to  
 CC target drugs to a tumour, in assays to determine biological activity, to  
 CC raise antibodies/ elicit an immune response, to determine quantitative  
 CC protein levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferative, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
 CC anti-inflammatory diseases, nervous system disorders, and infection.  
 CC The present sequence encodes a protein of the invention.

XX Sequence 3690 BP; 925 A; 947 C; 956 G; 862 T; 0 other;

Query Match 2.4%; Score 45.6; DB 22; Length 3690;  
 Best Local Similarity 49.2%; Pred. No. 0.011;  
 Matches 120; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1409 actttgatccagtagccgcgcatcaacacctggatctgttccactggtatcgcgcgacaca 1468

Db 1245 actttggcgtctcgttcaacctccagcactggtatctctcacaacaaagctggtcacct 1304

QY 1469 tcttgggtcaaggttgacaagatcaacatggcgaaactcccttgagctgcgagttccattct 1528

Db 1305 tgcctgtcagctttgtctcagctcaagaacctgaagtggttgacctgaagataaaccccc 1364

QY 1529 tggataaggagtttcaaggttcagagaccattcttacctgaagattgccaacg 1588

Db 1365 tggatcctgtcctggccaaggtggcaggtgactgctgttgatgagaagcagtgtaagcagt 1424

QY 1589 gtaccaccaagtagcgcgcgagggcactcgagcagattgttcgcctcagcttttgc 1648

Db 1425 gtgcaacaagaagttgttacagcacatgaaggccgtgcaggcagatcagcgaggagc 1484

QY 1649 accg 1652  
 Db 1485 agcg 1488  
 RESULT 15  
 AAC46946  
 ID AAC46946 standard; DNA; 1836 BP.  
 XX  
 AC AAC46946;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51990.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.

|    |              |               |    |              |               |
|----|--------------|---------------|----|--------------|---------------|
| PR | 17-JUN-1999; | 99US-0139492. | PR | 13-AUG-1999; | 99US-0148684. |
| PR | 18-JUN-1999; | 99US-0139454. | PR | 16-AUG-1999; | 99US-0149368. |
| PR | 18-JUN-1999; | 99US-0139455. | PR | 17-AUG-1999; | 99US-0149175. |
| PR | 18-JUN-1999; | 99US-0139456. | PR | 18-AUG-1999; | 99US-0149426. |
| PR | 18-JUN-1999; | 99US-0139457. | PR | 20-AUG-1999; | 99US-0149722. |
| PR | 18-JUN-1999; | 99US-0139458. | PR | 20-AUG-1999; | 99US-0149723. |
| PR | 18-JUN-1999; | 99US-0139459. | PR | 20-AUG-1999; | 99US-0149929. |
| PR | 18-JUN-1999; | 99US-0139460. | PR | 23-AUG-1999; | 99US-0149929. |
| PR | 18-JUN-1999; | 99US-0139461. | PR | 23-AUG-1999; | 99US-0149930. |
| PR | 18-JUN-1999; | 99US-0139462. | PR | 25-AUG-1999; | 99US-0150566. |
| PR | 18-JUN-1999; | 99US-0139463. | PR | 26-AUG-1999; | 99US-0150884. |
| PR | 18-JUN-1999; | 99US-0139750. | PR | 27-AUG-1999; | 99US-0151065. |
| PR | 18-JUN-1999; | 99US-0139763. | PR | 27-AUG-1999; | 99US-0151066. |
| PR | 21-JUN-1999; | 99US-0139817. | PR | 27-AUG-1999; | 99US-0151080. |
| PR | 22-JUN-1999; | 99US-0139899. | PR | 30-AUG-1999; | 99US-0151303. |
| PR | 23-JUN-1999; | 99US-0140353. | PR | 31-AUG-1999; | 99US-0151438. |
| PR | 23-JUN-1999; | 99US-0140354. | PR | 01-SEP-1999; | 99US-0151930. |
| PR | 24-JUN-1999; | 99US-0140695. | PR | 07-SEP-1999; | 99US-0152363. |
| PR | 28-JUN-1999; | 99US-0140823. | PR | 10-SEP-1999; | 99US-0153070. |
| PR | 29-JUN-1999; | 99US-0140991. | PR | 13-SEP-1999; | 99US-0153758. |
| PR | 30-JUN-1999; | 99US-0141287. | PR | 15-SEP-1999; | 99US-0154018. |
| PR | 01-JUL-1999; | 99US-0141842. | PR | 16-SEP-1999; | 99US-0154039. |
| PR | 01-JUL-1999; | 99US-0142154. | PR | 20-SEP-1999; | 99US-0154779. |
| PR | 02-JUL-1999; | 99US-0142055. | PR | 22-SEP-1999; | 99US-0155139. |
| PR | 06-JUL-1999; | 99US-0142390. | PR | 23-SEP-1999; | 99US-0155486. |
| PR | 08-JUL-1999; | 99US-0142803. | PR | 24-SEP-1999; | 99US-0155659. |
| PR | 09-JUL-1999; | 99US-0142920. | PR | 28-SEP-1999; | 99US-0156458. |
| PR | 12-JUL-1999; | 99US-0142977. | PR | 29-SEP-1999; | 99US-0156596. |
| PR | 13-JUL-1999; | 99US-0143342. | PR | 04-OCT-1999; | 99US-0157117. |
| PR | 14-JUL-1999; | 99US-0143624. | PR | 05-OCT-1999; | 99US-0157753. |
| PR | 15-JUL-1999; | 99US-0144005. | PR | 06-OCT-1999; | 99US-0157865. |
| PR | 16-JUL-1999; | 99US-0144085. | PR | 07-OCT-1999; | 99US-0158029. |
| PR | 16-JUL-1999; | 99US-0144086. | PR | 08-OCT-1999; | 99US-0158232. |
| PR | 18-JUL-1999; | 99US-0144325. | PR | 12-OCT-1999; | 99US-0158369. |
| PR | 18-JUL-1999; | 99US-0144331. | PR | 13-OCT-1999; | 99US-0158293. |
| PR | 19-JUL-1999; | 99US-0144332. | PR | 13-OCT-1999; | 99US-0158294. |
| PR | 19-JUL-1999; | 99US-0144333. | PR | 13-OCT-1999; | 99US-0159295. |
| PR | 19-JUL-1999; | 99US-0144334. | PR | 14-OCT-1999; | 99US-0159329. |
| PR | 19-JUL-1999; | 99US-0144335. | PR | 14-OCT-1999; | 99US-0159330. |
| PR | 20-JUL-1999; | 99US-0144352. | PR | 14-OCT-1999; | 99US-0159331. |
| PR | 20-JUL-1999; | 99US-0144632. | PR | 14-OCT-1999; | 99US-0159637. |
| PR | 20-JUL-1999; | 99US-0144684. | PR | 14-OCT-1999; | 99US-0159638. |
| PR | 21-JUL-1999; | 99US-0144814. | PR | 18-OCT-1999; | 99US-0159584. |
| PR | 21-JUL-1999; | 99US-0145086. | PR | 21-OCT-1999; | 99US-0160741. |
| PR | 21-JUL-1999; | 99US-0145088. | PR | 21-OCT-1999; | 99US-0160767. |
| PR | 22-JUL-1999; | 99US-0145085. | PR | 21-OCT-1999; | 99US-0160768. |
| PR | 22-JUL-1999; | 99US-0145087. | PR | 21-OCT-1999; | 99US-0160770. |
| PR | 22-JUL-1999; | 99US-0145089. | PR | 21-OCT-1999; | 99US-0160814. |
| PR | 22-JUL-1999; | 99US-0145192. | PR | 21-OCT-1999; | 99US-0160815. |
| PR | 23-JUL-1999; | 99US-0145145. | PR | 22-OCT-1999; | 99US-0160980. |
| PR | 23-JUL-1999; | 99US-0145218. | PR | 22-OCT-1999; | 99US-0160981. |
| PR | 23-JUL-1999; | 99US-0145224. | PR | 22-OCT-1999; | 99US-0160989. |
| PR | 26-JUL-1999; | 99US-0145276. | PR | 25-OCT-1999; | 99US-0161404. |
| PR | 27-JUL-1999; | 99US-0145913. | PR | 25-OCT-1999; | 99US-01       |





THIS PAGE BLANK (USPTO)  
THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 05:54:08 ; Search time 58.98 Seconds  
(without alignments)  
7996.208 Million cell updates/sec

Title: US-09-786-474-1

Perfect score: 1320

Sequence: 1 atgtggcgcttcttgat.....gtctaccggctgagctt 1920

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PCTUS-COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 116.6 | 6.1         | 30001  | 1 US-08-125-468-1   | Sequence 1, Appli |
| 2          | 116.6 | 6.1         | 30001  | 2 US-08-474-933-1   | Sequence 1, Appli |
| 3          | 37.6  | 2.0         | 1349   | 4 US-09-191-608-15  | Sequence 15, Appl |
| 4          | 37.6  | 2.0         | 1421   | 4 US-09-191-608-14  | Sequence 14, Appl |
| 5          | 37.6  | 2.0         | 1436   | 4 US-09-191-608-13  | Sequence 13, Appl |
| 6          | 37.6  | 2.0         | 1499   | 4 US-09-191-608-16  | Sequence 16, Appl |
| 7          | 36.4  | 1.9         | 1342   | 4 US-09-385-028-16  | Sequence 16, Appl |
| 8          | 36.4  | 1.9         | 11604  | 4 US-09-385-028-13  | Sequence 13, Appl |
| 9          | 36.4  | 1.9         | 15079  | 4 US-09-385-028-1   | Sequence 8, Appli |
| 10         | 36    | 1.9         | 531    | 4 US-09-191-608-8   | Sequence 1, Appli |
| 11         | 34.8  | 1.8         | 2267   | 4 US-08-679-645-25  | Sequence 25, Appl |
| 12         | 34.4  | 1.8         | 323    | 2 US-08-997-080-92  | Sequence 92, Appl |
| 13         | 34.4  | 1.8         | 323    | 2 US-08-997-362-92  | Sequence 92, Appl |
| 14         | 34.4  | 1.8         | 323    | 3 US-08-873-970-92  | Sequence 92, Appl |
| 15         | 34.4  | 1.8         | 323    | 4 US-09-095-855-92  | Sequence 92, Appl |
| 16         | 34.4  | 1.8         | 323    | 4 US-09-324-542-92  | Sequence 92, Appl |
| 17         | 34.4  | 1.8         | 1111   | 2 US-08-997-080-169 | Sequence 169, App |
| 18         | 34.4  | 1.8         | 1111   | 2 US-08-997-362-169 | Sequence 169, App |
| 19         | 34.4  | 1.8         | 1111   | 4 US-09-095-855-169 | Sequence 169, App |
| 20         | 34.4  | 1.8         | 1111   | 4 US-09-324-542-169 | Sequence 169, App |
| 21         | 34.4  | 1.8         | 1341   | 2 US-08-997-080-93  | Sequence 93, Appl |
| 22         | 34.4  | 1.8         | 1341   | 2 US-08-997-362-93  | Sequence 93, Appl |
| 23         | 34.4  | 1.8         | 1341   | 3 US-08-873-970-93  | Sequence 93, Appl |
| 24         | 34.4  | 1.8         | 1341   | 4 US-09-095-855-93  | Sequence 93, Appl |
| 25         | 34.4  | 1.8         | 1341   | 4 US-09-324-542-93  | Sequence 93, Appl |
| 26         | 34.4  | 1.8         | 2184   | 4 US-09-307-265A-2  | Sequence 2, Appli |
| 27         | 34.2  | 1.8         | 40138  | 3 US-09-090-793-12  | Sequence 12, Appl |

|   |    |      |     |         |   |                    |                    |
|---|----|------|-----|---------|---|--------------------|--------------------|
| c | 28 | 34.2 | 1.8 | 4411529 | 4 | US-09-103-840A-1   | Sequence 1, Appli  |
|   | 29 | 34   | 1.8 | 1050    | 4 | US-09-655-270A-16  | Sequence 16, Appl  |
|   | 30 | 34   | 1.8 | 1050    | 4 | US-09-651-941-20   | Sequence 20, Appl  |
|   | 31 | 34   | 1.8 | 1886    | 6 | 5256558-1          | Patent No. 5256558 |
|   | 32 | 34   | 1.8 | 12808   | 4 | US-09-655-270A-1   | Sequence 1, Appli  |
|   | 33 | 34   | 1.8 | 12523   | 4 | US-09-651-941-1    | Sequence 1, Appli  |
|   | 34 | 33.8 | 1.8 | 11459   | 4 | US-09-454-721A-3   | Sequence 3, Appli  |
|   | 35 | 33.6 | 1.8 | 3624    | 1 | US-08-434-730-13   | Sequence 13, Appl  |
| c | 36 | 33.6 | 1.8 | 4403765 | 4 | US-09-103-840A-2   | Sequence 2, Appli  |
|   | 37 | 33.4 | 1.7 | 2160    | 4 | US-09-588-256-1    | Sequence 1, Appli  |
|   | 38 | 33   | 1.7 | 1026    | 1 | US-07-975-526-6    | Sequence 6, Appli  |
|   | 39 | 33   | 1.7 | 1026    | 4 | US-07-974-409C-428 | Sequence 428, App  |
|   | 40 | 32.8 | 1.7 | 4689    | 4 | US-09-105-537-34   | Sequence 34, Appl  |
|   | 41 | 32.8 | 1.7 | 36778   | 4 | US-09-105-537-5    | Sequence 5, Appli  |
|   | 42 | 32.8 | 1.7 | 38506   | 3 | US-09-320-878-19   | Sequence 19, Appl  |
|   | 43 | 32.4 | 1.7 | 1220    | 3 | US-09-148-801-1    | Sequence 1, Appli  |
| c | 44 | 32.2 | 1.7 | 710     | 3 | US-09-098-287A-5   | Sequence 5, Appli  |
|   | 45 | 32.2 | 1.7 | 3301    | 2 | US-08-447-430A-42  | Sequence 42, Appl  |

ALIGNMENTS

RESULT 1  
US-08-125-468-1/c  
; Sequence 1, Application US/08125468  
; Patent No. 5589385  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmid  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125,468  
; FILING DATE: 22-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tsevdos, Estelle J  
; REGISTRATION NUMBER: 31,145  
; REFERENCE/DOCKET NUMBER: 31,255-02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201)831-3241  
; TELEFAX: (201)831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30001 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-125-468-1

Query Match 6.1%; Score 116.6; DB 1; Length 30001;  
Best Local Similarity 56.1%; Pred. NO. 1.4e-24;  
Matches 243; Conservative 0; Mismatches 184; Indels 6; Gaps 1;





```

; APPLICANT: Niforatos, Wende
; APPLICANT: Touma, Edward B.
; APPLICANT: Van Biesen, T.
; TITLE OF INVENTION: Nucleic Acids Encoding a Functional
; TITLE OF INVENTION: Human Purinoreceptor P2X2 And Methods Of Production
; TITLE OF INVENTION: And Use Thereof
; FILE REFERENCE: 6394.US.P1
; CURRENT APPLICATION NUMBER: US/09/191,608
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 1499
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-191-608-16

```

Query Match 2.0%; Score 37.6; DB 4; Length 1499;  
Best Local Similarity 47.4%; Pred. NO. 0.2;  
Matches 145; Conservative 0; Mismatches 159; Indels 2; Gaps 1;

```

RESULT 7
US-09-385-028-16
Sequence 16, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
;
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A Aidoo
; APPLICANT: Ashish S. Paradkar
;
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
;
; TITLE OF INVENTION: Acid Blosynthesis
;
; NUMBER OF SEQUENCES: 25
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PPLC
;
; STREET: The Jenifer Buliding, 400 Seventh Street, N.W.
;
; CITY: Washington
;
; STATE: D.C.
;
; COUNTRY: U.S.A.
;
; ZIP: 20004
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/385.028

```

FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/790,462  
 FILING DATE: 29-JAN-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: D. Douglas Price  
 REGISTRATION NUMBER: 24,514  
 REFERENCE/DOCKET NUMBER: 1418/P574520S2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 638-6666  
 TELEFAX: (202) 39305350  
 TELEX: RCA 248593 IDEA UR  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1542 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)

|    | Query Match           | 1.98;   | Score 36.4;     | DB 4;     | Length 1542; |
|----|-----------------------|---|-----------------|-----------|--------------|
|    | Best Local Similarity | 50.6;   | Pred. No. 0.47; |           |              |
|    | Matches 88;           | Conservative 0;   | Mismatches 86;  | Indels 0; | Gaps 0;      |
| Qy | 726                   | ggaicagagacctcttcgacgcattgcccaggtttggagagatagcgtcgcaaaagcatat | 785             |           |              |
| Db | 654                   | GGAGGGCGAGCCGTCGGGGCCGCTGGAGAGGCGCTGCGCCACGCGGT               | 713             |           |              |
| Qy | 786                   | gcgtgcgcagctgacctgtagctcgttcttccggcggccttgaactcaacgaattgc     | 845             |           |              |
| Db | 714                   | CACCCCGCGGCACACCCGTTGGTGGTCTCTCGGGGAAATCGACTCTCCGGGGTCCG      | 773             |           |              |
| Qy | 846                   | gccgcttgcgaagcgcacaaacctgaactgctcaccttcaccacgggtttcga         | 899             |           |              |
| Db | 774                   | GGCTGTGTGCGACCGCGCGCGGGGAACTGGACCGGTGTCCATGGGCACCGA           | 827             |           |              |

```

RESULT      8
US-09-385-028-13
; Sequence 13, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Keamena A Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Blosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN PPLC
; STREET: The Jenifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price

```

406 gaaagtcgcttttctctgcgcgtgatcagttcgc  
Qy  
354 GTCCAGGTCACAGTCCCAGTTGATATGACCCCG  
Db  
456 accgagcatggacccgtgtctctctcagagaaga  
Qy





Db 191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCGCAAGTTCACCGCGCCGTACATGACCG 250  
QY 506 tcttgagatggcggaggagatgaatctagatctggccttgataagcgaccattgagc 565  
Db 251 GCATGGTCGGTCTCGCTTACAACAAGCAGCCACCGGCGGATATCCGCAACCATGACG 310  
QY 566 actacgtggacc 577  
Db 311 ACCTCTGGGATC 322

RESULT 13  
US-08-997-362-92  
; Sequence 92, Application US/08997362  
; Patent No. 5985287  
; GENERAL INFORMATION:  
; APPLICANT: Tan, Paul  
; APPLICANT: Hiyama, Jun  
; APPLICANT: Visser, Elizabeth  
; APPLICANT: Skinner, Margot  
; APPLICANT: Scott, Linda  
; APPLICANT: Prestidge, Ross  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
; TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/997,362  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970  
; FILING DATE: June 12, 1997  
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347  
; FILING DATE: August 29, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sleath, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 11000.1002c2  
; TELEPHONE: 206-269-0565  
; TELEFAX: 206-269-0563  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 323 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-997-362-92

Query Match 1.8%; Score 34.4; DB 2; Length 323;  
Best Local Similarity 53.8%; Pred. No. 0.8;  
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 446 agccactgttctacgaaccaccagcatggcaccgtgttctctctcagagaagaccca 505  
Db 191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCGCAAGTTCACCGCGCCGTACATGACCG 250  
QY 506 tcttgagatggcggaggagatgaatctagatctggccttgataagcgaccattgagc 565  
Db 191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCGCAAGTTCACCGCGCCGTACATGACCG 250  
QY 506 tcttgagatggcggaggagatgaatctagatctggccttgataagcgaccattgagc 565

Db 251 GCATGGTCGGTCTCGCTTACAACAAGCAGCCACCGGCGGATATCCGCAACCATGACG 310  
QY 566 actacgtggacc 577  
Db 311 ACCTCTGGGATC 322

RESULT 14  
US-08-873-970-92  
; Sequence 92, Application US/08873970  
; Patent No. 6001361  
; GENERAL INFORMATION:  
; APPLICANT: Tan, Paul  
; APPLICANT: Hiyama, Jun  
; APPLICANT: Visser, Elizabeth  
; APPLICANT: Skinner, Margot  
; APPLICANT: Scott, Linda  
; APPLICANT: Prestidge, Ross  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
; TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/873,970  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/705,347  
; FILING DATE: 29-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sleath, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 11000.1002C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-269-0565  
; TELEFAX: 206-269-0563  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 323 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-873-970-92

Query Match 1.8%; Score 34.4; DB 3; Length 323;  
Best Local Similarity 53.8%; Pred. No. 0.8;  
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 446 agccactgttctacgaaccaccagcatggcaccgtgttctctctcagagaagaccca 505  
Db 191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCGCAAGTTCACCGCGCCGTACATGACCG 250  
QY 506 tcttgagatggcggaggagatgaatctagatctggccttgataagcgaccattgagc 565  
Db 251 GCATGGTCGGTCTCGCTTACAACAAGCAGCCACCGGCGGATATCCGCAACCATGACG 310  
QY 566 actacgtggacc 577

Search completed: September 20, 2002, 07:14:05  
Job time: 4797 sec

|    | Query Match           | 1.88;   | Score 34.4;    | DB 4; | Length 323;                      |
|----|-----------------------|---|----------------|-------|----------------------------------|
|    | Best Local Similarity | 53.88;  | Pred. NO. 0.8; |       |                                  |
|    | Matches               | 71;   | Conservative   | 0;    | Mismatches 61; Indels 0; Gaps 0; |
| QY | 446                   | agccactgttctacgcaaccacgagcatggacacggtgttctctctcagagaagaccca | 505            |       |                                  |
|    |                       |   |                |       |                                  |
| Db | 191                   | AGGACCTGTTGGATCTCAGGATTCAGCAGGCGCCCAAGTTACCGCGCGTACATGACCG  | 250            |       |                                  |
|    |                       |   |                |       |                                  |
| QY | 506                   | tcttggagatgcccggaggatgaattagatcttggccttgaagaagcaccattgagc   | 565            |       |                                  |
|    |                       |   |                |       |                                  |
| Db | 251                   | GCATGGTCGGTCTCGCCCTACACAAGCAGCCACCGGCGGATATCGGACCATTGACG    | 310            |       |                                  |
|    |                       |   |                |       |                                  |
| QY | 566                   | actacgtggacc  | 577            |       |                                  |
|    |                       |   |                |       |                                  |
| Db | 311                   | ACCTCTGGGATC  | 322            |       |                                  |

---

**THIS PAGE BLANK (USPTO)**

| Result No. | Score | Query |       | Length | DB | ID        | Description         |
|------------|-------|-------|-------|--------|----|-----------|---------------------|
|            |       | Match | Match |        |    |           |                     |
| C          | 1     | 58.6  | 3.1   | 635    | 12 | BH0011105 | BH0011105 G179 Gemm |
|            | 2     | 56.2  | 2.9   | 552    | 9  | AA697638  | AA697638 HL02916.5  |
|            | 3     | 56.2  | 2.9   | 565    | 9  | A1516024  | A1516024 LD41758.5  |
|            | 4     | 56.2  | 2.9   | 575    | 9  | A1387646  | A1387646 GH18225.5  |
|            | 5     | 56.2  | 2.9   | 601    | 10 | B1630083  | B1630083 RH59048.5  |
|            | 6     | 56.2  | 2.9   | 602    | 10 | B1239433  | B1239433 RE35917.5  |
|            | 7     | 56.2  | 2.9   | 609    | 9  | A1062991  | A1062991 GH02388.5  |
|            | 8     | 56.2  | 2.9   | 616    | 10 | B1641347  | B1641347 SD24579.5  |
|            | 9     | 56.2  | 2.9   | 626    | 10 | B1212840  | B1212840 RE18085.5  |
|            | 10    | 56.2  | 2.9   | 630    | 9  | A1389247  | A1389247 GH20374.5  |
|            | 11    | 56.2  | 2.9   | 644    | 10 | B1635460  | B1635460 SD16825.5  |
|            | 12    | 56.2  | 2.9   | 693    | 10 | B1142084  | B1142084 SD16082.5  |
|            | 13    | 56.2  | 2.9   | 701    | 10 | B1641950  | B1641950 SD25636.5  |
|            | 14    | 55    | 2.9   | 640    | 10 | B1719579  | B1719579 I03104480  |
|            | 15    | 54.6  | 2.8   | 620    | 10 | B1354992  | B1354992 GM29151.5  |
|            | 16    | 53.6  | 2.8   | 601    | 10 | B1527119  | B1527119 I02407800  |
|            | 17    | 53.6  | 2.8   | 664    | 10 | B1874100  | B1874100 963113009  |



|                           |  |  |                 |
|---------------------------|--|--|-----------------|
| Qy                        | 294  | ctttaatactctctgcatgagcccaattgttctggtttccaccactgggagtc            | 353             |
| Db                        | 381  | CTACAACCCCATGAGCGATTGCCACGTGATACACTGGAACGTGACCAGGATTATGGAAGGA    | 440             |
| Qy                        | 354  | cgtagtgagcatctcgcggaatgtctgcgcgcattgcattgggatacaaaaggaaagtc      | 413             |
| Db                        | 441  | TCATTCTGCAATACATCACTACGGAATGTTTGGCTTTTGGCTTGTACGATAGGAAGACCAAGGA | 500             |
| Qy                        | 414  | gctttcttcctgcgctgcatcagttcgcgcacaaagccactgt                      | 454             |
| Db                        | 501  | AGTCTCTTCTGCCGAGACCCCTTTGGCATTTATACCCATGT                        | 541             |
| RESULT                    | 5  |  |                 |
| Bi630083                  | Bi630083   | 601 bp   | linear          |
| LOCUS                     | RH59048.5  | prime RH Drosophila melanogaster                                 | EST 10-SEP-2001 |
| DEFINITION                | Drosophila melanogaster  | cdna clone RH59048.5, mRNA sequence.                             |                 |
| ACCESSION                 | Bi630083   |  |                 |
| VERSION                   | Bi630083.1   | GI:15532293  |                 |
| KEYWORDS                  | EST.   |  |                 |
| SOURCE                    | fruit fly.   |  |                 |
| ORGANISM                  | Drosophila melanogaster  |  |                 |
| REFERENCE                 | 1 (bases 1 to 601)   |  |                 |
| AUTHORS                   | Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C. J., Nuno, J., Pacle, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.   |  |                 |
| TITLE                     | BDGP/HMI RH Drosophila EST Project   |  |                 |
| JOURNAL                   | Unpublished (2001)   |  |                 |
| COMMENT                   | Contact: Stapleton, M. BDGP Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: <a href="http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu">http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu</a> hit genomic AE003593; arm:3L [20799555;21112669] estimated-cyto:78A2-78C7: 08/23/2001 Plate: RH.590 row: D column: 12 High quality sequence stop: 578. Location/Qualifiers |  |                 |
| FEATURES                  | source   |  |                 |
|                           | 1..601   |  |                 |
|                           | /organism="Drosophila melanogaster"  |  |                 |
|                           | /db_xref="taxon:7227"  |  |                 |
|                           | /clone="RH59048"   |  |                 |
|                           | /clone_lib="RH Drosophila melanogaster normalized head pFlc-1"   |  |                 |
|                           | /sex="male and female"   |  |                 |
|                           | /dev_stage="Adult"   |  |                 |
|                           | /lab_host="DH5-alpha Tona"   |  |                 |
|                           | /note="Organ: head; Vector: pFlc1; Site:1; xhoi; BamHI: Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cdna library."  |  |                 |
| BASE COUNT                | 157 a 130 c 164 g 150 t  |  |                 |
| ORIGIN                    |  |  |                 |
| Query Match               | 2.9%   | Score 56.2; DB 10; Length 601;                                   |                 |
| Best Local Similarity     | 53.4%  | Pred. No. 0.00049;   |                 |
| Matches 118; Conservative | 0; Mismatches 103; Indels  | 0; Gaps  | 0;              |
| Qy                        | 234  | caacggtgagatctacaactacgttcgtagctgcgtaaaagctctcggattgggataac      | 293             |
| Db                        | 275  | CAATGGAGAGATCTACAATTATCTCGAGCTTTCGGCTGAGATTGCGAAGACCGTGGCTC      | 334             |
| Qy                        | 294  | ctttaatactctgcgtagcgagcccaatgtctgcgcatttcaccactgggagtc           | 353             |
| Db                        | 335  | CTACAACCCCATGAGCGATTGCCACGTGATACACTGGAACGTGACCAGGATTATGGAAGGA    | 394             |
| Qy                        | 354  | cgtagtgagcatctcgcggaatgtctgcgcacaaagccactgt                      | 413             |
| Db                        | 395  | TCATTCTGCAATACATCACTGGAATGTTTGGCTTTTGGCTTGTACGATAGGAAGACCAAGGA   | 454             |
| Qy                        | 414  | gctttcttcctgcgctgcatcagttcgcgcacaaagccactgt                      | 454             |
| Db                        | 455  | AGTCTCTTCTGCCGAGACCCCTTTGGCATTTATACCCATGT                        | 495             |
| RESULT                    | 4  |  |                 |
| Al387646                  | Al387646   | 575 bp   | linear          |
| LOCUS                     | GH18225.5  | prime GH Drosophila melanogaster                                 | EST 19-APR-2001 |
| DEFINITION                | melanogaster cdna clone GH18225.5  | prime, mRNA sequence.  |                 |
| ACCESSION                 | Al387646   |  |                 |
| VERSION                   | Al387646.1   | GI:4201657   |                 |
| KEYWORDS                  | EST.   |  |                 |
| SOURCE                    | fruit fly.   |  |                 |
| ORGANISM                  | Drosophila melanogaster  |  |                 |
| REFERENCE                 | 1 (bases 1 to 575)   |  |                 |
| AUTHORS                   | Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G. M.   |  |                 |
| TITLE                     | BDGP/HMI Drosophila EST Project  |  |                 |
| JOURNAL                   | Unpublished (2001)   |  |                 |
| COMMENT                   | Contact: Stapleton, M. BDGP Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: <a href="http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu">http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu</a> Plate: 182 row: C column: 1 High quality sequence stop: 508. Location/Qualifiers   |  |                 |
| FEATURES                  | source   |  |                 |
|                           | 1..575   |  |                 |
|                           | /organism="Drosophila melanogaster"  |  |                 |
|                           | /db_xref="taxon:7227"  |  |                 |
|                           | /clone="GH18225"   |  |                 |
|                           | /clone_lib="GH Drosophila melanogaster head pOr2"  |  |                 |
|                           | /sex="male and female"   |  |                 |
|                           | /dev_stage="adult"   |  |                 |
|                           | /lab_host="DH5 - alpha"  |  |                 |
|                           | /note="Organ: head; Vector: pOr2; Site:1; EcoRI; Site:2; xhoi; SmaI; SmaI fractionated cDNAs were directly ligated into pOr2. Plasmid cdna library."   |  |                 |
| BASE COUNT                | 154 a 124 c 158 g 139 t  |  |                 |
| ORIGIN                    |  |  |                 |
| Query Match               | 2.9%   | Score 56.2; DB 9; Length 575;                                    |                 |
| Best Local Similarity     | 53.4%  | Pred. No. 0.00048;   |                 |
| Matches 118; Conservative | 0; Mismatches 103; Indels  | 0; Gaps  | 0;              |
| Qy                        | 234  | caacggtgagatctacaactacgttcgtagctgcgtaaaagctctcggattgggataac      | 293             |
| Db                        | 321  | CAATGGAGAGATCTACAATTATCTCGAGCTTTCGGCTGAGATTGCGAAGACCGTGGCTC      | 380             |

```

Db 401 CTACAACCCCATGACCGATTGCCACGGTGATCTAGGAAGTATGGAAGGATTTATGGAAGGA 460
Qy 354 cgtggtcagcatctccgcgggaattgttcggcattgccatttgggatacaaaaggaaaagtc 413
Db 461 TCCTCTGCAATACATCACTAGTGAATGTTTGGCTTTGACGATAGGAGACCAAGGA 520
Qy 414 gcttttccttcggtgatcagtttcggcatcaagccactgt 454
Db 521 AGTGCTTCTTCCCGAGACCCCTTTGGCATTATATACCCATGT 561

RESULT 6
BI239433 602 bp mRNA linear EST 12-JUL-2001
LOCUS RE35917.5prime RE Drosophila melanogaster normalized Embryo pF1c-1
DEFINITION Drosophila melanogaster cDNA clone RE35917 5, mRNA sequence.
ACCESSION BI239433
VERSION BI239433.1 GI:14707974
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 602)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
,G.M.
BDGP/HHMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003593: arm:3L (20799555,21112669]
estimated-cyto:78A2-78C7: 05/12/2001
Plate: RE.359 row: B column: 5
High quality sequence stop: 508.
Location/Qualifiers
FEATURES
source
1. .602
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE35917"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pF1c-1"
/dev_stage="male and female"
/note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 158 a 129 c 166 g 149 t
ORIGIN
Query Match 2.9%; Score 56.2; DB 10; Length 602;
Best Local Similarity 53.4%; Pred. No. 0.0049;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 234 caacggtgagatctacaactcagttgagctcgtaaaagactctcggtttgggatacac 293
Db 340 CAATGGAGAGATCTACAATTTTGGCTTTGCGTGAGATTGCGAAAAGCGTGGCTC 399
Qy 294 ctttaatactctcggcagcgagcgaattgttcggtttccaccactgggcgagtc 353
Db 400 CTACAACCCCATGACCGATTGCCACGGTGATCTGGAAGTATGGAAGGA 459

Query Match 2.9%; Score 56.2; DB 10; Length 602;
Best Local Similarity 53.4%; Pred. No. 0.0049;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 234 caacggtgagatctacaactcagttgagctcgtaaaagactctcggtttgggatacac 293
Db 340 CAATGGAGAGATCTACAATTTTGGCTTTGCGTGAGATTGCGAAAAGCGTGGCTC 399
Qy 294 ctttaatactctcggcagcgagcgaattgttcggtttccaccactgggcgagtc 353
Db 400 CTACAACCCCATGACCGATTGCCACGGTGATCTGGAAGTATGGAAGGA 459

```

```

Qy 354 cgtggtcagcatctccgcgggaattgttcggcattgccatttgggatacaaaaggaaaagtc 413
Db 460 TCCTCTGCAATACATCACTAGTGAATGTTTGGCTTTGACGATAGGAGACCAAGGA 519
Qy 414 gcttttccttcggtgatcagtttcggcatcaagccactgt 454
Db 520 AGTGCTTCTTCCCGAGACCCATTTGGCATTATATACCCATGT 560

RESULT 7
AI062991 609 bp mRNA linear EST 19-APR-2001
LOCUS GH02388.5prime GH Drosophila melanogaster head pOT2 Drosophila
DEFINITION melanogaster cDNA clone GH02388 5prime, mRNA sequence.
ACCESSION AI062991
VERSION AI062991.1 GI:3338830
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 609)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 23 row: H column: 4
High quality sequence stop: 564.
Location/Qualifiers
FEATURES
source
1. .609
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH02388"
/clone_lib="GH Drosophila melanogaster head pOT2"
/dev_stage="male and female"
/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT 146 a 145 c 167 g 151 t
ORIGIN
Query Match 2.9%; Score 56.2; DB 9; Length 609;
Best Local Similarity 53.4%; Pred. No. 0.0005;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 234 caacggtgagatctacaactcagttgagctcgtaaaagactctcggtttgggatacac 293
Db 12 CAATGGAGAGATCTACAATTTTGGCTTTGCGTGAGATTGCGAAAAGCGTGGCTC 71
Qy 294 ctttaatactctcggcagcgagcgaattgttcggtttccaccactgggcgagtc 353
Db 72 CTACAACCCCATGACCGATTGCCACGGTGATCTGGAAGTATGGAAGGA 131
Qy 354 cgtggtcagcatctccgcgggaattgttcggcatccatttgggatacaaaaggaaaagtc 413
Db 132 TCCTCTGCAATACATCACTAGTGAATGTTTGGCTTTGACGATAGGAGACCAAGGA 191
Qy 414 gcttttccttcggtgatcagtttcggcatcaagccactgt 454
Db 192 AGTGCTTCTTCCCGAGACCCCTTTGGCATTATATACCCATGT 232

RESULT 8

```



|                       |   |                                  |                           |                |                 |
|-----------------------|---|----------------------------------|---------------------------|----------------|-----------------|
| BI641347              | BI641347  | 616 bp                           | mRNA                      | linear         | EST 10-SEP-2001 |
| LOCUS                 | SD24579.5   | prime SD Drosophila melanogaster | Schneider L2 cell culture |                |                 |
| DEFINITION            | pot2  | Drosophila melanogaster          | cDNA clone SD24579 5,     | mRNA sequence. |                 |
| ACCESSION             | BI641347  |                                  |                           |                |                 |
| VERSION               | BI641347.1  | GI:15543557                      |                           |                |                 |
| KEYWORDS              | EST.  |                                  |                           |                |                 |
| SOURCE                | fruit fly.  |                                  |                           |                |                 |
| ORGANISM              | Drosophila melanogaster   |                                  |                           |                |                 |
| REFERENCE             | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |                                  |                           |                |                 |
| AUTHORS               | 1 (bases 1 to 616)  |                                  |                           |                |                 |
|                       | Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.   |                                  |                           |                |                 |
| TITLE                 | BDGP/HHMI Drosophila EST Project  |                                  |                           |                |                 |
| JOURNAL               | Unpublished (2001)  |                                  |                           |                |                 |
| MENT                  | Contact: Stapleton, M.  |                                  |                           |                |                 |
|                       | BDGP  |                                  |                           |                |                 |
|                       | Lawrence Berkeley National Lab  |                                  |                           |                |                 |
|                       | One Cyclotron Rd, Berkeley, CA 94720, USA   |                                  |                           |                |                 |
|                       | Fax: 510 486 6798   |                                  |                           |                |                 |
|                       | Email: <a href="http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu">http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu</a>  |                                  |                           |                |                 |
|                       | hit genomic AE003593; arm:3L [20799555,21112669]  |                                  |                           |                |                 |
|                       | estimated-cyto:78A2-78C7: 05/19/2001  |                                  |                           |                |                 |
|                       | Plate: SD.245 row: G column: 7  |                                  |                           |                |                 |
|                       | High quality sequence stop: 511.  |                                  |                           |                |                 |
| FEATURES              | Location/Qualifiers   |                                  |                           |                |                 |
| source                | 1..616  |                                  |                           |                |                 |
|                       | /organism="Drosophila melanogaster"   |                                  |                           |                |                 |
|                       | /db_xref="taxon:7227"   |                                  |                           |                |                 |
|                       | /clone="SD24579"  |                                  |                           |                |                 |
|                       | /clone_11b="SD Drosophila melanogaster Schneider L2 cell culture pot2"  |                                  |                           |                |                 |
|                       | /lab_host="DH5-alpha"   |                                  |                           |                |                 |
|                       | /note="vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."  |                                  |                           |                |                 |
| BASE COUNT            | 161 a 133 g 173 g 149 t   |                                  |                           |                |                 |
| ORIGIN                |   |                                  |                           |                |                 |
| Query Match           | 2.9%; Score 56.2; DB 10; Length 616;  |                                  |                           |                |                 |
| Best Local Similarity | 53.4%; Pred. No. 0.0005;  |                                  |                           |                |                 |
| Matches               | 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  |                                  |                           |                |                 |
| Db                    | 234 caacggtgagatctcacactacgttgagtcgcgtaaagagctctcgattgggataac 293   |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| Db                    | 316 CAATGGAGAGATCTACAAATATCTCGGAGCTTCGGCTGAGATTCGAAAAAGCGTGCGTC 375   |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| QY                    | 294 ctttaatactctcgcatggcagaccaaattgttgcgtttccaccactggggcgagtc 353   |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| Db                    | 376 CTACAACCCCATGAGCGAATGCCACGTGATCTGGAAGTGTACCCAGGATTTATGGAAGA 435   |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| QY                    | 354 cgtggtcgagcatctcgcggaattgttcgcatgccatttgggatacaagaagaaaagtc 413   |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| Db                    | 436 TCCTTCGAATACATCACTGGAATGTTTGTCTTTTGCCTGTACGATAGGAAGACCAAGGA 495   |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| QY                    | 414 gcttttccttcgctgctgatacagttcgcgcatcaagccactgt 454  |                                  |                           |                |                 |
|                       |   |                                  |                           |                |                 |
| Db                    | 496 AGTCTCTCTGCGCGAGACCCCTTTGCGCATATATACCCATGT 536  |                                  |                           |                |                 |

| RESULT     | 9   |
|------------|---|
| BI212840   |   |
| LOCUS      | BI212840 626 bp mRNA linear EST 11-JUL-2001   |
| DEFINITION | RE18085.5prime RE Drosophila melanogaster normalized Embryo pf1c-1 Drosophila melanogaster cDNA clone RE18085 5, mRNA sequence. |
| ACCESSION  | BI212840  |
| VERSION    | BI212840.1 GI:14690555  |
| KEYWORDS   | EST.  |
| SOURCE     | fruit fly.  |
| ORGANISM   | Drosophila melanogaster   |

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,p., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouaneavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.

TITLE BDGP/HHMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu  
hit genomic AE003593; arm:3L [20799555,21112669]  
estimated-cyto:78A2-78C7: 04/12/2001  
Plate: RE.180 row: H column: 1  
High quality sequence stop: 592.

FEATURES  
source Location/Qualifiers  
1..626  
    /organism="Drosophila melanogaster"  
    /db\_xref="taxon:7227"  
    /clone="RE18085"  
    /clone\_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"  
    /sex="male and female"  
    /dev\_stage="0-24 hours mixed stage embryonic"  
    /lab\_host="DH5-alpha Tona"  
    /note="Organ: embryo; Vector: pFlc1; Site\_1: XhoI; Site\_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

BASE COUNT     162 a   134 c   175 g   155 t

ORIGIN

Query Match               2.9%   Score 56.2; DB 10; Length 636;  
Best Local Similarity   53.4%; Pred.No.0.0005;  
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 234 caacggtgagatctacaactcgtagctgcgtaaaagagctctcgattgggatataac 293  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 340 CAATGGAGAGATCACAAATTATCTGGAGCTTTCGCGTGAGATTCCGAATAACGCTGCCTC 399  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 294 cttaataactctctggcgatggcgagccaatttgcttgcggttcaccacctggggcgagtcc 353  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 400 CTACAACCCCATGAGCGATTGCCAGTTCCTCATCTACTGGAACGTACCAGGATTATGGAAGAAG 459  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 354 cgtggtcgagcatctcgcggaatgttcggcatcgccatttgggatatacaaggaaaagtc 413  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 460 TCCTTCTGCAATACATCACTTGAATGTTTTCCTTTTGCCTTGACGATAGGAAGAACCAAGGA 519  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 414 gctttccttcgcggtgatcagcttcggcatcaagccactgt 454  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 520 AGTCTCTTGCCGAGACCCCTTTGGCATTTATACCCATGT 560  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10

AI389247

LOCUS AI389247 630 bp mRNA linear EST 19-APR-2001

DEFINITION GH20374.5 prime GH Drosophila melanogaster head pot2 Drosophila melanogaster cDNA clone GH20374 5prime, mRNA sequence.

ACCESSION AI389247

VERSION AI389247.1 GI:4203258

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera.

```
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 630)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 203 row: G column: 2
High quality sequence stop: 408.

FEATURES
source
1..630
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH20374"
/clone_lib="GH Drosophila melanogaster head pOT2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DHS - alpha"
/notes="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT 166 a 138 c 179 g 146 t 1 others
ORIGIN

Query Match 2.9%; Score 56.2; DB 9; Length 630;
Best Local Similarity 53.4%; Pred. No. 0.0005;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcgatttgggatatac 293
Db 330 CAATGGAGAGATCTACAATTTGAGCTTCGCGCTGAGATTGCGAAAAACGCGTGGCTC 389

Qy 294 ctttaatactctcgagatcgagcgaattgttcggtttccaccactgggcgagtc 353
Db 390 CTACAACCCCATGAGCGATTGCCAGCTGATCTGGAACCTGTACCAAGATTATGGAAGA 449

Qy 354 cgtggtcgagatctccgcgaattgttcggtttccaccactgggcgagtc 413
Db 450 TCTTCTGCAATACATCACTGGAATGTTTGCCTTTGCCCTGTACGATAGGAAGCAAGGA 509

Qy 414 gctttcttcgctgagtcagttcgccatcaagccactgt 454
Db 510 AGTGTCTTCTGCCGAGACCCCTTTGGCATATATACCCATGT 550

RESULT 11
LOCUS
DEFINITION
SD16825.Prime SD Drosophila melanogaster Schneider L2 cell culture
pOT2 Drosophila melanogaster cDNA clone SD16825 5, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 644)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 203 row: G column: 2
High quality sequence stop: 408.

FEATURES
source
1..630
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH20374"
/clone_lib="GH Drosophila melanogaster head pOT2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DHS - alpha"
/notes="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT 166 a 138 c 179 g 146 t 1 others
ORIGIN

Query Match 2.9%; Score 56.2; DB 10; Length 644;
Best Local Similarity 53.4%; Pred. No. 0.00051;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcgatttgggatatac 293
Db 314 CAATGGAGAGATCTACAATTTGAGCTTCGCGCTGAGATTGCGAAAAACGCGTGGCTC 373

Qy 294 ctttaatactctcgagatcgagcgaattgttcggtttccaccactgggcgagtc 353
Db 374 CTACAACCCCATGAGCGATTGCCAGCTGATCTGGAACCTGTACCAAGATTATGGAAGA 433

Qy 354 cgtggtcgagatctccgcgaattgttcggtttccaccactgggcgagtc 413
Db 434 TCTTCTGCAATACATCACTGGAATGTTTGCCTTTGCCCTGTACGATAGGAAGCAAGGA 493

Qy 414 gctttcttcgctgagtcagttcgccatcaagccactgt 454
Db 494 AGTGTCTTCTGCCGAGACCCCTTTGGCATATATACCCATGT 534

RESULT 12
LOCUS
DEFINITION
SD16825.Prime SD Drosophila melanogaster Schneider L2 cell culture
pOT2 Drosophila melanogaster cDNA clone SD16825 5, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 693)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 203 row: G column: 2
High quality sequence stop: 604.

FEATURES
source
1..693
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/lab_host="DHS-alpha"
/notes="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 166 a 143 c 177 g 158 t
ORIGIN
```

```

/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/lab_host="DH5-alpha"
/notes="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT      177 a   150 c   194 g   172 t
ORIGIN

Query Match      2.9%; Score 56.2; DB 10; Length 693;
Best Local Similarity 53.4%; Pred. No. 0.00053;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 234 caacgtgagatcacaactagctgagctgtaagagctctcgattgggataac 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 CAATGGAGAGATCTACAAATTTATCTGGAGCTTTCGGCTGAGATTGCCGAAAGCGTGCTC 375

QY 294 cttaatactctcgagcgagcgaattgtgtcgtttccaccactggggcgagtc 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CTACAACCCCATGAGCGATTGCCAGTGTACTGGAACTGTACCAGGATTTATGGAAGGA 435

QY 354 cgtggtcgagcatctccggaattgttcgcatctggcattgggatacaaaagaaagtc 413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 TCTTCTGCAATACATCCTGGAATGTTTGTCTTTTGGCTTGTACGATAGGAAGCAAGGA 495

QY 414 gctttccctggcgatcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 AGTGCTTCTTGGCCGAGACCCCTTTGGCATTTATACCATGT 536

RESULT 13
BI641950
LOCUS      701 bp mRNA linear EST 10-SEP-2001
DEFINITION SD25636.Sprime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD25636 5, mRNA sequence.
ACCESSION BI641950
VERSION    BI641950.1 GI:15544160
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 701)
AUTHORS    Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
            hit genomic AC00593; arm:3L [20799555,21112669]
            estimated-cyto:78A2-78C7: 05/19/2001
            Plate: SD.256 row: C column: 12
            High quality sequence stop: 574.

FEATURES
            Location/Qualifiers
            1..701
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="SD25636"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pOT2"
            /lab_host="DH5-alpha"
            /notes="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
            fractionated cDNAs were directly ligated into pOT2.
            Plasmid cDNA library."
BASE COUNT      182 a   151 c   195 g   172 t   1 others

```

```

ORIGIN

Query Match      2.9%; Score 56.2; DB 10; Length 701;
Best Local Similarity 53.4%; Pred. No. 0.00053;
Matches 118; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 234 caacgtgagatcacaactagctgagctgtaagagctctcgattgggataac 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 CAATGGAGAGATCTACAAATTTATCTGGAGCTTTCGGCTGAGATTGCCGAAAGCGTGCTC 380

QY 294 cttaatactctcgagcgagcgaattgtgtcgtttccaccactggggcgagtc 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 CTACAACCCCATGAGCGATTGCCAGCTGTACTGGAACCTGTACCAGGATTTATGGAAGGA 440

QY 354 cgtggtcgagcatctccggaattgttcgcatctggcattgggatacaaaagaaagtc 413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 TCTTCTGCAATACATCCTGGAATGTTTGTCTTTTGGCTTGTACGATAGGAAGCAAGGA 500

QY 414 gctttccctggcgatcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 AGTGCTTCTTGGCCGAGACCCCTTTGGCATTTATACCATGT 541

RESULT 14
BI719579
LOCUS      640 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031044E03.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI719579
VERSION    BI719579.1 GI:15695274
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
ORGANISM   Chlamydomonas reinhardtii.
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadales; Chlamydomonas.
REFERENCE 1 (bases 1 to 640)
AUTHORS    Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
JOURNAL    Unpublished (2001)
COMMENT    Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.

FEATURES
            Location/Qualifiers
            1..640
            /organism="Chlamydomonas reinhardtii"
            /strain="CC-1690 wild type mt+ 21gr"
            /db_xref="taxon:3055"
            /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
            ), Lambda Zap II"
            /notes="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
            XhoI; Stress condition II library, constructed by John
            Davies and Jeffrey McDermott, combines cDNAs from CC-1690
            cells grown to mid-log phase in TAP (NH4+ - containing)
            and shifted to TAP - NO3- (24hrs); H2 production
            conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
            Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
            sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
            PolyA mRNA was purified from each sample, pooled and cDNA
            synthesized. The cDNA was directionally cloned into lambda
            Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
            sites. pBluescript II SK- plasmids were excised from the
            lambda Zap clones by superinfection with EXASist
            (Stratagene) phage. The library was normalized using
            method 4 described in Bonaldo et al., (1996) Genome
            Research 6: 791-806."

```

| BASE COUNT |  | 129 a | 201 c | 185 g | 125 t | FEATURES |  | Location/Qualifiers  |  |
|------------|--|-------|-------|-------|-------|----------|--|--|--|
| ORIGIN     |  |       |       |       |       | source   |  | 1. 620   |  |
|            |  |       |       |       |       |          |  | /organism="Drosophila melanogaster"  |  |
|            |  |       |       |       |       |          |  | /db_xref="taxon:7227"  |  |
|            |  |       |       |       |       |          |  | /clone_lib="GM29151"   |  |
|            |  |       |       |       |       |          |  | /clone_lib="GM Drosophila melanogaster ovary pOT2"   |  |
|            |  |       |       |       |       |          |  | /sex="female"  |  |
|            |  |       |       |       |       |          |  | /dev_stage="newly eclosed females: germarium-stage 6"  |  |
|            |  |       |       |       |       |          |  | /lab_host="XLI Blue"   |  |
|            |  |       |       |       |       |          |  | /note="Organ: ovary; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2." |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |
|            |  |       |       |       |       |          |  |  |  |

---

**THIS PAGE BLANK (USPTO)**



1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17  
1 ATGTGGGGCTTCTTGGCATATTGACTGCAATGGGAACGCTCAAGCAT 50  
17 eValProAlaLeuGluAlaLeuProCysMetArgHisArgGlyProA 34  
51 CGTTCGACCTCGAGCGGGCTTGCATGCGCCACCGTGGTCCTG 100  
34 sPAspAlaGlyThrTrpHisAspAlaAspAlaPheGlyPheAsnArg 50  
101 ACATCGCGGCACTTGGCATGACCGCGATGCACGTTTGATTCACCGC 150  
51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67  
151 CTCTCCATCATGATATGCACACTCCCAACCACTGCGTTGGGGACC 200  
67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleYrA 84  
201 TGGGGATGAACCGACCGCTACGCAATGACTTTCAACGGTGAGATCAGA 250  
84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100  
251 ACTACGTTGAGCTGCGTAAGAGCTCTCGAATTTGGGATATACCTTTAAT 300  
101 ThrSerGlyAspGlyGluProIleValAlGlyPheHisHisTrpGlyGI 117  
301 ACTTCTGGCGATGGCGAGCAATTTGTGCGTTTCCACCACCTGGGGCGA 350  
117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspTr 134  
351 GTCCGTGGTGGAGCATCTCCGCGAATGTTGGGCATTTGCCATTTGGGATA 400  
134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150  
401 CAAGGAAAAGTCGCTTTCCCTTGGCGGTGATCAGTTCCGCAATCAAGCCA 450  
151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGlyLysLy 167  
451 CTGTTCTAGCAACCAACCGAGCATGGCACCGTGTCTCTCAGAGAAGAA 500  
167 sThrIleLeuGluMetAlaGluMetAsnLeuAspLeuGlyLeuAspL 184  
501 GACCATTTGGAGATGGCGAGAGATGAATCTAGATCTGGGCGCTTGATA 550  
184 ysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProAsp 200  
551 AGCGCAACCATTGAGCACTACGTGGACCTGCAGTACGTGCCGAGCCAGAT 600  
201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217  
601 ACCCTTCACGCGCAGATTTCCCGCTTGGAGTCAGGCTGCACCGCAACAGT 650  
217 lArgProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheP 234  
651 TCGTCGGGGCGGCAAGCTGGAACAGAAGGTTACTTCAAGCCCTCAGTTCC 700  
234 roValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIle 250  
701 CAGTACAGAAGTTCGTAAGGTTAAGGACGAGGACGACCTCTTCGATCGCAT 750  
251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267  
751 GCCCAGGTGTTGGAGTACGCGTGAAGAGCATATGCGTGCCGACGTGAC 800  
267 rValGlySerPheLeuSerGlyIleAspSerThrAlaIleAlaProl 284  
801 CGTAGGCTGCTTCTTCCGGGGGATTAAGTCAACCGCAATTTGCCCGCG 850  
284 euAlaLysArgHisAsnProAspLeuLeuThrPheThrThrGlyPheGlu 300  
851 TTGCAAGGCCACACCCCTGACCTCCTCACCTTCACCGGTTTCGAG 900

301 ArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaAlaIle 317  
901 CGTGAAGGCTACTCGGAGGTGATGTGGTGGGAGTCCGCGCTCGCAT 950  
317 eGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsnA 334  
951 TGGCGCTGAGCACATCGTGAAGATTGTCGCTGAGGAATACGCCAAGC 1000  
334 laIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer 350  
1001 CGATTCCTAAGATCATGTGTACTTGGATGATCCTGTAGCTGACCCATCA 1050  
351 LeuValProLeuTyrPheValAlaAlaGluAlaArgLysHisValLysVa 367  
1051 TTGGTCCCGCTGTACTTCGTGGCAGCGAAGCAGTAAAGCAGCTCAAGGT 1100  
367 lValLeuSerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleT 384  
1101 TGTGCTGTCTGGCGAGGGCGAGATGAGCTGTTGCGTGGATACCATTT 1150  
384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu 400  
1151 ACNAGAGCGGCTATCGCTTGCCTTGCATTTGAGAAAGATCCCTCCCCACTA 1200  
401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGI 417  
1201 CGTAAAGGCTGGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGG 1250  
417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG 434  
1251 CAAGTCCCTTCTTGAGCGTGGCTCCATGACCATGGAAGAGCGCTACTACG 1300  
434 lYasnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp 450  
1301 GCNAGCGCTCGCTCCTTCATTTGAGCAGATGCACGCGTTATTCATGG 1350  
451 AlaLysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGI 467  
1351 GCNAGCGGAAATGGGACCAACCGGAAGTCACTGCACCGCATCAGCACA 1400  
467 nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT 484  
1401 ATCCCGCAACTTTGATCCAGTAGCCCGCATGCAACACCTGGGATCTGTTC 1450  
484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500  
1451 CCTGGATGCGCGGACATCTTGGTCAAGGCTGACAAGATCAACATGGCG 1500  
501 AsnSerLeuGluLeuArgValProPheLeuAspLysGluValPheLysVa 517  
1501 AACTCCCTTGAGCTGCGAGTTCCATTTCTTGGATAAGGAAGTTTCAAGGT 1550  
517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT 534  
1551 TGCAGAGACCATTTCTTACGATCTGAAGATTGCCAACGGTACCACCAAGT 1600  
534 yrAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHis 550  
1601 ACGCGCTGCGCAGGGCACTCGAGCAGATTGTTCCGCTCACGTTTGCAC 1650  
551 ArgLysLysLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs 567  
1651 CGCAGAAGCTGGGCTTCCCTGTTCCATCGCCACTGGCTTGCCTGCGCGCA 1700  
567 pGluLeuPheGlyTrpAlaGlnAspThrIleLysGluSerGlyThrGluA 584  
1701 TGAGCTGTTCCGTTGGCGCAGGACACCATTAAGGAATCCGGTACTGAAG 1750  
584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgAsp 600  
1751 ATATCTTCAACAAGCAGGCTGTGCTGGATATGCTGAACGAGCACCGCAT 1800  
601 GlyValSerAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa 617



```

|||||
1801 GCGGTGTCAGATCATTCCTCGACTGTGGACTGTCTGTCTATTTATGGT 1850
617 ltrPhisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634
1851 GTGGCAGCGCATTTTGTGGAAACCGCATTCATTCACAGATTGAGGACC 1900
634 rgSerTyrProValGluLeu 640
1901 GCTCCTACCGGTCGAGCTT 1920

```

seq\_name: /STD1/gcdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA10293

seq\_documentation\_block:

ID AAA10293 standard; DNA; 3825 BP.

XX AC AAA10293;

XX 03-JUL-2000 (first entry)

DE DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.

XX Lysozyme insensitivity protein; bacterium; recombinant expression;

XX amino acid production; ds.

XX Corynebacterium glutamicum.

XX Key Location/Qualifiers

XX CDS 815..2737

XX /tag= a  
/product= "Corynebacterium glutamicum lysozyme  
insensitivity protein"

XX WO200014241-A1

XX 16-MAR-2000.

XX 04-SEP-1998; 98WO-JP03981.

XX 04-SEP-1998; 98WO-JP03981.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nagai K, Wati M;

XX WPI; 2000-256989/22.

XX P-PSDB; AAY87459.

XX New DNA molecule useful for production of amino acids encodes protein  
impacting lysozyme resistance to Corynebacterium glutamicum strain -

XX Example 2; Page 40-44; 50pp; Japanese.

XX This sequence represents DNA encoding a Corynebacterium glutamicum  
lysozyme insensitivity protein. This protein renders lysozyme-sensitive  
strains of Corynebacterium glutamicum insensitive to lysozyme. DNA  
sequences encoding the lysozyme insensitivity protein can be used to  
construct a recombinant vector for the expression of this protein in a  
host cell. The DNA encoding the lysozyme insensitivity protein can also  
be mutated, and used to generate strains of Corynebacterium glutamicum in  
which this protein is inactivated and which are therefore lysozyme-  
sensitive. Corynebacterium glutamicum is used for the production of amino  
acids, in particular glutamic acid and glutamine.

XX Sequence 3825 BP; 881 A; 1031 C; 993 G; 920 T; 0 other;

alignment\_scores:

Quality: 3379.00 Length: 640  
Ratio: 5.280 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

```

US-09-786-474-2 x AAA10293 ..
Align seg 1/1 to: AAA10293 from: 1 to: 3825

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaph 17
815 ATGTGCGGCTTCTTTGGCATATTGACTGCAATGGAACGCTGAAGCATT 864

17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
865 GGTTCCTGCATCTCGAGCGGCCCTTCCTCATGTCATGCCACCGTGTCTCTG 914

34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg 50
915 ACGATGCCGCGCATTGGCATGACGCCGATGCAGCGTTTGGATTCAACCGC 964

51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67
965 CTCTCCATCATTTGATATTCACACTCCCAACCAACCACTGCGTTGGGACC 1014

67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA 84
1015 TCGCGATGACCCGACCGCTAGCGCAATGACTTCAACGGTGAGATCTACA 1064

84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
1065 ACTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTCGGATATACCTTTAAT 1114

101 ThrSerGlyAspGlyGluProIleValValGlyPheHisHisTrpGlyGl 117
1115 ACTTCTGGCGATGGCGGCCAATTTGTGTGGTTTCCACCACTGGGGCGA 1164

117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
1165 GTCCGTGTTCGAGCATCTCCGCGGAATGTTCCGCAATGCCATTTGGGATA 1214

134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150
1215 CAAAGGAAAGTCGCTTTTCTTTCGCGGTGATCAGTTCGGCATCAAGCCA 1264

151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLysLy 167
1265 CTGTTCTACGCAACCAACCGAGCATGGCAGCTGCTCTCTCAGAGAAGAA 1314

167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspL 184
1315 GACCATCTGGAGATGGCGGAGAGATGAACTAGATCTGGCCCTTGATA 1364

184 ysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProasp 200
1365 AGCGCACCATTTGAGCAGCTACGTGGACCTGCAGTACGTGCCCGAGCAGAT 1414

201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217
1415 ACCCTTCACGGCAGATTTCCGCTTGGAGTCAGGCTGCACCGCAACAGT 1464

217 lArgProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheP 234
1465 TCGTCCGCGGCGCAAGCTGGACACAGACGCTTACTTCAAGCCTCAGTTCC 1514

234 roValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgile 250
1515 CAGTACAGAAGGTCCTAAAGGGTAAAGGAGCAGGACCTCTCTCGATCCGATT 1564

251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267
1565 GCCCAGGTGTTGGAGGATAGCGTCGAAAAGCATATGCGTGGCGGAGTGAC 1614

267 rValGlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProL 284
1615 CGTAGGCTCGTTCTTTCGCGCGGATGACTCAACCGCAATTGGCGCGC 1664

284 euAlaLysArgHisAsnProAspLeuLeuThrPheThrThrGlyPheGlu 300

```

|||||  
1665 TTCAAAGCCCAACCTGACCTGCTACCTTCCACCGGTTTCGAG 1714  
301 ArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaAlaI 317  
|||||  
1715 CGTGAAGCTACTCGGAGGTGATGCTGCTGCGGAGTCCCGCTCGAT 1764  
317 eGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsnA 334  
|||||  
1765 TGGCGCTGACACATCGTGAAGATTGCTCGCTGAGGAATAGCCCAACG 1814  
334 laileProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer 350  
|||||  
1815 CGATTCTTAAGATCATGTGTGCTACTTGGATGATCTGTAGTACCCATCA 1864  
351 LeuValProLeuTyrPheValAlaAlaGluAlaArgLysHisValLysVa 367  
|||||  
1865 TTGGTCCCGCTGTACTTCTGCGGAGCGGAGACGTAAGCACCATCAAGT 1914  
367 lValLeuSerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIle 384  
|||||  
1915 TGTGCTGTCTGCGAGGCGCACATGAGCTGTTCGGTGGATACACCATTT 1964  
384 yrlYsGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu 400  
|||||  
1965 ACAAGAGCGCGCTATCGCTTGTCTTCCATTTGAGAAGATCCCTCCCACTA 2014  
401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysG1 417  
|||||  
2015 CGTAAAGCCCTGGGAAGCTCACAGAGTTCTGCCAGACGGCATGAAGG 2064  
417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTrg 434  
|||||  
2065 CAAGTCCCTTCTTGGCGGTGGCTCCATGACCATGGAAGACGCTACTACG 2114  
434 lyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp 450  
|||||  
2115 GCAAGCGTCTGCTCTTCAATTTTCAGCAGATGCAACGCGTTATTCATGG 2164  
451 AlAlysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaG1 467  
|||||  
2165 GCAAAGCGCAATGGGACCCAGCGGAAGTCACCTGCACCGATCTACGCACA 2214  
467 nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT 484  
|||||  
2215 ATCCCGCAACTTTGATCCAGTAGCCGCGATGCAACACCTGGATCTGTCA 2264  
484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500  
|||||  
2265 CTTGGATGCGCGGACATCTCTGCTCAAGCTGACAAGATCAACATGGCG 2314  
501 AsnSerLeuGluLeuArgValProPheLeuAspLysGluValPheLysVa 517  
|||||  
2315 AACTCCTTGGAGTGGCGAGTTCCATCTTGGATAAGGAAGTTTCAAGGT 2364  
517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT 534  
2365 TGCAGAGACATTTCTTACCATCTGAAGATGCGCAACGCTACCAACCAAGT 2414  
534 yrAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHis 550  
|||||  
2415 ACGGCTGCGCAGGCGACTGACAGACATTTGCTCCGCTCAGCTTTTGCAC 2464  
551 ArgLysLysLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs 567  
|||||  
2465 CGCAAGAAGTGGGCTTCCCTGTTCCTCATGCGCCACTGGCTTCCGCGCA 2514  
567 pGluLeuPheGlyTrpAlaGlnAspThrIleLysGluSerGlyThrGluA 584  
2515 TGAGCTGTTCGGTGGCGCAGACACCATTAAGGAATCCGGTACTGAAG 2564  
584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgAsp 600  
|||||

2565 ATATCTTCAACAAGAGGCTGTGCTGATATGCTGAACGAGCACCAGCAT 2614  
601 GlyValSerAspHisSerArgArgLeuTyrThrValLeuSerPheMetVa 617  
|||||  
2615 GGGCTGTCAATCATTTCCGCTCGACTGTGACTGTTCTGTTCATTTATGGT 2664  
617 lTrpHisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634  
|||||  
2665 GTGGCAGCGCATTTTGTGAAAAACCGCATTTGATCCACAGATTGAGGACC 2714  
634 rgSerTyrProValGluLeu 640  
|||||  
2715 GCTCTACCGGTCGAGCTT 2734

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAH67371

seq\_documentation\_block:

ID AAH67371 standard; DNA; 1920 BP.

XX AC AAH67371;

XX XX 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 2406.

XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN BPI-08790-A2.

XX PD 20-JUN-2001.

XX XX 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX DR P-PSDB; AAG92152.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying

XX PT mutation point of a gene, measuring expression of a gene, analysing

XX PT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 8; SEQ ID NO: 2406; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein

XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

XX CC are useful for identifying the mutation point of a gene derived from a

XX CC mutant of coryneform bacterium, measuring expression amount and

XX CC analysing the expression profile or expression pattern of a gene derived

XX CC from Coryneform bacterium, and identifying a homologue of a gene derived

XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

XX CC particularly L-lysine. The present sequence is a nucleic acid described

XX CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from the

XX CC European Patent Office.

XX SQ Sequence 1920 BP; 422 A; 534 C; 528 G; 436 T; 0 other;

alignment\_scores:

Quality: 3366.00 Length: 640  
Ratio: 5.268 Gaps: 0  
Percent Similarity: 99.844 Percent Identity: 99.688  
alignment\_block:  
US-09-786-474-2 x AAH67371 ..  
Align seg 1/1 to: AAH67371 from: 1 to: 1920  
1 MetCysGlyLeuLeuGlyIleLeuThraAlaAsnGlyAsnAlaGluAlaph 17  
1 ATGTGGCGCCTTCTTGGCATATTGACTGCAAAATGGGAACGCTGAAGCAATT 50  
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34  
51 CGTTCTCTGACATCGAGCGGGCTTGCCATGTCATGCGCACCGTGGTCCTG 100  
34 sPAspAlaGlyThrTrpHisaspAlaAlaPheGlyPheAsnArg 50  
101 ACGATGGCGGCACATTGGGCATGACGCCGATGACGGCTTTGGATTCAACCGC 150  
51 LeuSerIleLeAspIleAlaHisSerHisGlnProLeuArgTTPGlyPr 67  
151 CTCCTCCATCATGTATTTGCACATCTCCACCACCACTGCTGGTGGGGACC 200  
67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA 84  
201 TGGCGATGAACCCGACCGCTACGCAATGACTTTC AACGGTGAGATCTACA 250  
84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100  
251 ACTACGTTGAGCTCGTAAAGAGCTCTCGGATTTGGGATATTCCTTTAAT 300  
101 ThrSerGlyAspGlyGluProIleValGlyPheHisHisTTPGlyG1 117  
301 ACTTCTGGCGATGGCGACCAATTGTTGCGGTTTCCACCACTGGGGCGA 350  
117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTTPAspT 134  
351 GTCCGTGGTCGAGCATCTCCCGGGAATGTCGGCATTTGCCATTTGGGATA 400  
134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150  
401 CAAGGAAAGTCGCTTTTCCTTCGCGTGATCAGTTTCGGCATTTAAGCCA 450  
151 LeuPheTyrAlaThrGluHisGlyThrValPheSerSerGluLysLy 167  
451 CTGTTCTACGCAACACCGACATGGCACCGTGTTCCTCAGAGAAGAA 500  
167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspL 184  
501 GACCATCTTGAGATGGCGGAGGAGATGAATCTAGATCTGGGCTTGATA 550  
184 ysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProasp 200  
551 AGCCACCATTTGACACTACGTGGACTTGCAATGTCAGTCCGCCGAGCCAGAT 600  
201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217  
601 ACCCTTACGGCGCAGATTTCCTGCTTGGAGTTCAGGCTGCACCGCAACAGT 650  
217 lArgProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheP 234  
651 TCGTCCGGCGCAAGCTGGAACAGAGACGTTACTTCAAGCCTCAGTTCC 700  
234 roValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIle 250  
701 CAGTACAGAGGCTGTAAGGGTAAGGAGCAGGACCTCTTCGATCGCATT 750  
251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267  
751 GCCCAGGTTGGAGGATAGCGTCGAAAGCATATGCGTCCCGCAGCGTAC 800

267 rValGlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProL 284  
801 CGTAGGCTCGTTCCCTTCGGCGGCATTGACTCAACCGCAATTGGCGGC 850  
284 euAlaLysArgHisAsnProAspLeuLeuThrPheThrThrGlyPheGlu 300  
851 TTGCAAAAGCGCCACAACTTACCTGCTCAGCTTCACCCCGGTTTCGAG 900  
301 ArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaAlaAlaI 317  
901 CGTGAAGGCTACTCGGAGTGCATGTGGCTCGGAGTCCGGCGCTGCCAT 950  
317 eGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsnA 334  
951 TGGCGCTGAGCACATCGTGAAGATTGCTCGCTGAGGAATACGCCAAG 1000  
334 laIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer 350  
1001 CGATTCTCTAAGATCATGTGGTACTTTGGATGATCTCTAGCTGACCCATCA 1050  
351 LeuValProLeuTyrPheValAlaAlaGluAlaArgLysHisValLysVa 367  
1051 TTGTTCCCGCTGTACTTCTGGCAGCGGAGACGTAAGCAGCTCAAGT 1100  
367 lValLeuSerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleT 384  
1101 TGTGCTGTCTGGCAGGCGCAGATGAGCTGTTCGGTGGATACACCATTT 1150  
384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu 400  
1151 ACAAGGAGCCGCTATCGCTTCTCCATTTGAGAAGATCCCTTCCCACATA 1200  
401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysG1 417  
1201 CGTAAAGCCCTGGGAAAGCTCAGCAAGGTTCTGCCAGCAGCGCATGAGGG 1250  
417 yLysSerLeuLeuGluArgGlySerMetThrMetGluArgTyrTyrG 434  
1251 CAAGTCCCTTCTTGAGCTGTGCTCCATGACCATGGAAGAGCGCTACTACG 1300  
434 lYasnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp 450  
1301 GCAACGCTCGCTCCTTCAATTTTCGAGCAGATGCCAACGCGTTATTCCATGG 1350  
451 AlaLysArgGluTTPAspHisArgGluValThrAlaProIleTyrAlaG1 467  
1351 GCAAGCGCGAATGGGACCAACCGGAGTCACTGCGCGCATCTACGCACA 1400  
467 nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT 484  
1401 GTCCCGCAACTTTGATCCAGTAGCCCGCATGCAACACCTGGATCTGTTC 1450  
484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500  
1451 CCTGGATCGCGGCGACATCCTGTGTCGAAGGCTGACAAGATCAACATGGCG 1500  
501 AsnSerLeuGluLeuArgValProPheLeuAspLysGluValPheLysVa 517  
1501 AACTCCCTTGAGCTGCGAGTTCCTATCTTGGATAAGGAAGTTTTCAGGT 1550  
517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT 534  
1551 TGCAGAGACCATTCCTTACGACCTGAAGATTGCCAACGGTACCACCAAGT 1600  
534 yrAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHis 550  
1601 ACGCGCTGCGCAGGCGACTCGAGCAGATTGTTCCGCTCAGCTTTTGCAC 1650  
551 ArgLysLysLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs 567  
1651 GCAGAAAGCTGGGCTTCCCTGTTCCCATGCGCCACTGGCTTGGCGGGA 1700  
567 pGluLeuPheGlyTrpAlaGlnAspThrIleLysGluSerGlyThrGluA 584

|||||  
1701 TGAGCTGTTGGTGGCGGAGACACCATCAAGGAATCCGGTACTCAAG 1750  
XX  
584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgAsp 600  
XX  
1751 ATATCTCAACAAGCAGGCTGCTGGATGCTGAACGAGACCGCGAT 1800  
PI  
601 GlyValSerAspHisSerArgLeuTrpThrValLeuSerPheMetVa 617  
|||||  
1801 GGCCTGCAGATCATCCCGTCGACTGTGGACTGTTCTGTCTATATGGT 1850  
XX  
617 lTrpHisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634  
|||||  
1851 GTGGCAGCGCATTTTGTGGAAACCGCATGTCATGCACAGATTGAGGACC 1900  
634 rgSerTyrProValGluLeu 640  
|||||  
1901 GCTCCTACCCAGTCGAGCTT 1920

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH68531

documentation\_block:  
AAH68531 standard; DNA; 349980 BP.

AC AAH68531;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 7066.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic-acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

PN 20-JUN-2001.

XX 18-DEC-2000;--2000EP-0127688.

PD 16-DEC-1999; 95JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280989.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

PI WPI; 2001-376931/40.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT -

XX Disclosure; SEQ ID NO: 7066; 246pp + Sequence Listing; English.

PS The present invention provides a number of nucleotide and protein

XX sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of corynebacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Corynebacterium bacterium, and identifying a homologue of a gene derived

CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX

SQ Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;  
alignment\_scores:  
Quality: 3366.00 Length: 640  
Ratio: 5.268 Gaps: 0  
Percent Similarity: 99.844 Percent Identity: 99.688  
alignment\_block:  
US-09-786-474-2 x AAH68531 ..

Align seg 1/1 to: AAH68531 from: 1 to: 349980

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17  
228516 ATGTGGGGCTTCTTGGCATATTGACTGCAAAATGGGACGCTGAAGCAATT 228565  
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34  
228566 CGTTCCTGCACTCGAGGGGCTTGGCATGCATGCGCCACCGTGGTCTG 228615  
34 spAspAlaGlyThrTrpHisAspAlaAspAlaPheGlyPheAsnArg 50  
228616 ACATGCCGGCACTTGGCATGACGCCGATGCAGCGTTTGGATTCAACCGC 228665  
51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67  
228666 CTCTCCATCATTTGATATTCACACTCCCAACCACTCGTGGGGACC 228715  
67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA 84  
228716 TCGGATGAACCGACCGCTACGCAATGACTTTCAACGGTGAGATCTACA 228765  
84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100  
228766 ACTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTGGGATATGCTTTAAT 228815  
101 ThrSerGlyAspGlyGluProIleValGlyPheHisHisTrpGlyGl 117  
228816 ACTTCTGGCGATGGCGAGGCAATTTGTTCGGTTTCCACCACTGGGGCGA 228865  
117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134  
228866 GTCCGTGGTGAGCATCTCCGGGAATTTTCGGCATTTGCCATTTGGGATA 228915  
134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150  
228916 CAAAGGAAAGTCGCTTTTCCITTCGCGTGATCAGTTCCGCAATTAAGCCA 228965  
151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLysLy 167  
228966 CTGTTCTAGCAACCAACCGAGCATGGCACCGTGTCTCTCTCAGAGAAGAA 229015  
167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspL 184  
229016 GACCATCTTGGAGATGGCGAGGAGATGAATCTAGATCTGGGGCTTGATA 229065  
184 ysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProAsp 200  
229066 AGCGCACCATTTGAGCACTACGTGGGACTTCCAGTACGTGCGCCGAGCCAGAT 229115  
201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217  
229116 ACCCTTCACGCGCAGATTTCCTGGAGTTCAGGCTGAGGCTGAGGCAACAGT 229165  
217 largProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheP 234  
229166 TCGTCCGGCGCGCAAGCTGGAACAGAGCGTTACTTCAAGCCCTCAGTTCC 229215  
234 roValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIle 250  
229216 CAGTACAGAGAGTCTGAAGGGTAAGGAGCAGGACCTCTTCCATCGCAATT 229265

251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267  
229266 GCCCAGGTTGGAGGATAGCTCGAAAGCATATGCGTCCGACGTGAC 229315  
267 rValGlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProL 284  
229316 CGTAGGCTCGTTCCTTCCGGCGCATTTGACTCAACCGCAATTCGGCGC 229365  
284 euAlaLysArgHisAsnProAspLeuLeuThrPheThrThrGlyPheGlu 300  
229366 TTGCANAGCGCCACACCTGACCTGCTACCTTCACACCGGTTTCGAG 229415  
301 ArgGluGlyTyrSerGlnValAspValAlaAlaGluSerAlaAlaIle 317  
229416 CGTGAAGGCTACTCGGAGGTCCATGTGTGCTCGGAGTCCGCGCTCGCAT 229465  
317 eGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsnA 334  
229466 TGGCGCTGAGCAGCATCGTGAAGATTGTCTCGCCTGAGGAATACGCCAACG 229515  
334 laIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer 350  
229516 CGATTCTTAAGATCATGTGGTATGATGCTGTAGCTGACCCATCA 229565  
351 LeuValProLeuTyrPheValAlaAlaGluAlaArgLysHisValLysVa 367  
229566 TTGGTCCGCTGTACTTCTGTCGACGCGAAGCAGCATAGCAGCTCAAGGT 229615  
367 lValLeuSerGlyGluGlyAlaAspGluLeuPheGlyTyrThrIleT 384  
229616 TGTGCTGTCTGGCAGGCGCAGATGAGCTGTCTGGTGGATACACCATTT 229665  
384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu 400  
229666 ACAAGGAGCCGCTATTCGCTTCTCATTGTGAGAGATCCCTTCCCACTA 229715  
401 ArgLysGlyLeuGlyLysSerLysValLeuProAspGlyMetLysG1 417  
229716 CGTAAAGCCCTGGGAAAGCTCAGCAAGTTCTGCCAGACGCGCATGAAGG 229765  
417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG 434  
229766 CAAGTCCCTTCTTGAGCTGCTCCATGACCATGGAGAGCGCTACTACG 229815  
434 lYAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp 450  
229816 GCAACGCTCGCTCCTCAATTCGAGCAGATGCAACGCGTTATTCATGG 229865  
451 AlAtysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaG1 467  
229866 GCAAGCGCGAATGGGACCCAGCGAAGTCACTGGCGCGATCTAGGCACA 229915  
467 nSerArgAsnPheAspProValAlaArgMetGlnHisIleAspLeuPheT 484  
229916 GTCCCGCAACTTGTATCCAGTAGCCCGCATGCAACACCTGGATCTGTCA 229965  
484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500  
229966 CCTGGATCGCGGCGCACATCTGTCAAGGTGACCAAGATCAACATGGCG 230015  
501 AsnSerLeuGluLeuArgValProPheLeuAspLysGluValPheLysVa 517  
230016 AACTCCCTTGAGCTCGAGTTCCTTCATTCTTGGATAAGGAAGTTTCAAGT 230065  
517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrLysT 534  
230066 TGCAGAGACCATTCCTTACGACCTGAAGATTGCCAACGGTACCCAAAGT 230115  
534 yrAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHis 550  
230116 ACGCCGCTCGGAGGCGACTCGAGCAGATGTTCCGCCCTCAGCTTTGCAC 230165  
551 ArgLysLysLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs 567

230166 CGCAAGAAGCTGGGCTTCCCTGTTCCCATGCGCCACTGCTTGGCGGCA 230215  
567 pGluLeuPheGlyTrpAlaGlnAspThrIleLysGluSerGlyThrGluA 584  
230216 TGAGCTGTTCGGTTGGGCGCAGGACACCATCAAGGAATCCGGTACTGAAG 230265  
584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgAsp 600  
230266 ATATCTTCAACACAGCAGGCTGTGCTGGATATGCTGAACGAGCAGCGCAT 230315  
601 GlyValSerAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa 617  
230316 GCGGTGTGAGATCATTCCTCGTACGCTGTGGACTGTCTGCTCATTTATGTT 230365  
617 lTrpHisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634  
230366 GTGGCAGGCGATTTTGTGGAAACCGCATTCATCCACAGATTGAGGACC 230415  
634 rgSerTyrProValGluLeu 640  
230416 GCTCCTACCCAGTCGAGCTT 230435  
seq\_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF71807  
seq\_documentation\_block:  
ID AAF71807 standard; DNA; 1962 BP.  
XX  
AC AAF71807;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:109.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
XX  
OS Cořynebacterium glutamicum.  
XX  
PN WO200100843-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF --23-JUN-2000, 2000WO-IB00923.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.

PR 09-JUL-1999; 99DE-1032130.  
 PR 09-JUL-1999; 99DE-1032186.  
 PR 09-JUL-1999; 99DE-1032206.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032228.  
 PR 09-JUL-1999; 99DE-1032229.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032926.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1033004.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 12-AUG-1999; 99US-0148613.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041380.  
 PR 31-AUG-1999; 99DE-1041394.  
 PR 03-SEP-1999; 99DE-1041396.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042124.  
 PR 09-MAR-2000; 2000US-0187970.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M., Kröeger B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 DR WPI: 2001-137957/14.  
 DR P-PSDB; AAB79688.  
 XX  
 XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
 PT pathway proteins, useful for producing fine chemicals in  
 PT microorganisms, including organic acids, nonproteinogenic amino acids,  
 PT and purine and pyrimidine bases -  
 XX  
 PS Claim 3; Page 324-327; 1737pp; English.  
 XX  
 CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
 CC MP nucleic acids are useful for the production of fine chemicals  
 CC in microorganisms, including organic acids, nonproteinogenic amino  
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
 CC compounds, vitamins, cofactors, polyketides and enzymes.  
 XX  
 SQ Sequence 1962 BP; 435 A; 544 C; 534 G; 449 T; 0 other:

alignment\_scores:  
 Quality: 3360.00 Length: 640  
 Ratio: 5.266 Gaps: 0  
 Percent SImilarity: 99.688 Percent Identity: 99.531

alignment\_block:  
 US-09-786-474-2 x AAF71807 ..  
 Align seg 1/1 to: AAF71807 from: 1 to: 1962

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaph 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 20 ATGTGCGGCTTCTTGGCATATTGACTGCAATGGGACGCTGAACATT 69

17 eValProAlaLeuGluAraGAlaLeuProCysMetArgHisArgGlyProA 34  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 70 CGTTCCTGCACTCGAGCGGGCCTTGCCATGCATGCGCCACCGTGGCTTG 119  
 34 sPAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg 50  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 120 ACATGCGGCGCATTGGCATGCGCGGATGCAGCGTTTGGATTCAACCGC 169  
 51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuAraGlyPr 67  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 170 CTCCTCATCATTTGATATTGCACACTCCCAACCACTCGGTGGGGACC 219  
 67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA 84  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 220 TGGGGATGAACCGACCGCTAGCAATGACTTTCAACGGTGAGATCTACA 269  
 84 sNtYrValGluLeuAraGlyLysGluLeuSerAspLeuGlyTyrThrPheAsn 100  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 270 ACTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTGGGATATGCTTTAAT 319  
 101 ThrSerGlyAspGlyGluProIleValGlyPheHisHisTrpGlyG 117  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 320 ACTTCTGGCGATGGCGAGCAATTTGTTCGGTTTCCACCACCTGGGGCGA 369  
 117 uSerValValGluHisLeuAraGlyMetPheGlyIleAlaIleTrpAspT 134  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 370 GTCCGTGGTGGAGCATCTCCGGGAATGTTCCGCATTGCCATTGGGATA 419  
 134 hrlYsGluLysSerLeuPheLeuAlaAraGlnPheGlyIleLysPro 150  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 420 CAAAGGAAAGTCGCTTTTCCTTTCGCGGTGATCAGTTCGGCATTAAGCCA 469  
 151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLysLy 167  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 470 CTGTTCTAGCAACACCGAGCATGGCACCCTGTTCTCTCAGAGAAGAA 519  
 167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspL 184  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 520 GACCATCTTGGAGATGGCGAGGAGATGAATCTAGATCTGGGCTTGATA 569  
 184 ySArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProAsp 200  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 570 AGCGCACCATATTGAGCATACGTGGACTTGAGTACGTGCCCGAGCCAGAT 619  
 201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 620 ACCCTTCACGCGCAGATTTCCTCCGCTTGGAGTCAGGCTGCACCGCAACAGT 669  
 217 lArgProGlyGlyLeuGluGlnLysArgTyrPheLysProGlnPheP 234  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 670 TCGTCCGGGGCGGCAAGCTGGAACAGAAAGCGTTACTTCAAGCCTCAGTTCC 719  
 234 roValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIle 250  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 720 CAGTACAGAAGTCGTAAAGGGTAAAGGACGAGCACCTCTTCGATCGCAT 769  
 251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 770 GCCCAGGTGTTGGAGGATAGCGTCGAAAGACATATCGGTGCCGACGTGAC 819  
 267 rValGlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProL 284  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 820 CGTAGGCTCGTTCTTTTCGGGGCATTGACTCAACCGCAATTCGGCGCG 869  
 284 euAlaLysArgHisAsnProAspLeuLeuThrPheThrThrGlyPheGlu 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 870 TTGCAAAAGCGCCACAACCTGACCTGCTCACCTTACCACCGGTTTCGAG 919  
 301 ArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaAlaAlaI 317  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 920 CGTGAAGGCTACTCGGAGGTGCGATGTGGTGGGAGTCCCGCGCTCGAT 969  
 317 eGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsnA 334

|||||  
970 TGGCGCTCAGCACATCGTGAAGATTGCTCGCTCAGGAATACGCCAACG 1019  
|||||  
334 laileProLysileMetTrpTyrLeuAspAspProValalaAspProSer 350  
|||||  
1020 CGATTCTTAAGATCATGTGTACTTGGATGATCTGTAGCTGACCCATCA 1069  
|||||  
351 LeuValProLeuTyrPheValalaAlaGluAlaArgLysHisValLysVa 367  
|||||  
1070 TTGGTCCGCTGTACTTCTGGCAGCGAAGCAGTAAAGCAGCTCAAGGT 1119  
|||||  
367 lvalLeuSerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleT 384  
|||||  
1120 TGTGCTGTCTGGCAGGCGCAGATGAGCTGTTCGGTGATACACCATTT 1169  
|||||  
384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu 400  
|||||  
1170 ACAAGGAGCCGCTATCGTTCCTCCATTTGAGAGATCCCTCCCACTA 1219  
|||||  
401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysG 417  
|||||  
1220 CGTAAAGGCTTGGGAAAGCTCAGCAAGGTTCTGCCAGACGCGATGAAGG 1269  
|||||  
417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG 434  
|||||  
1270 CAAGTCCCTCTTGGAGCTGTGCTCCATGACCATGGAGAGCGCTACTACG 1319  
|||||  
434 lyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp 450  
|||||  
1320 GCACGCTCGCTCCTCAATTTCCAGCAGATGCAACGGTATTTCATGG 1369  
|||||  
451 AlaLysArgGluTTrpAspHisArgGluValThrAlaProIleTyrAlaG 467  
|||||  
1370 GCAAGCGCAATGGGACCGACCGGAAGTCACGTGCGCGATCTACGCACA 1419  
|||||  
467 nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT 484  
|||||  
1420 GTCCGCAACTTTGATCCAGTAGCCCGATGCACACCTGGATCTGTTCAC 1469  
|||||  
484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500  
|||||  
1470 CTGTGATCGCGGCGACATCTGTGTCAAGGCTGACAAAGATCAACATGCG 1519  
|||||  
501 AsnSerLeuGluLeuArgValProPheLeuAspLysGluValPheLysVa 517  
|||||  
1520 AACTCCCTTGAGCTGCGAGTTCATCTTGGATAAGGAAGTTTCAAGGT 1569  
|||||  
517 lalaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT 534  
|||||  
1570 TGCAGAGACCATCTCTTACGACCTGAAGATTGCCAACGGTACCACCAAGT 1619  
|||||  
534 yrAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHis 550  
|||||  
1620 ACGCGCTCGCAGGCGCACTCGAGCAGATTGTTCCGCCCTCACGTTTGCAC 1669  
|||||  
551 ArgLysLysLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs 567  
|||||  
1670 CGCAAGAAGCTGGGCTTCCCTGTTCCTCCATGCGCCACTGGCTTGGCGCGA 1719  
|||||  
567 pGluLeuPheGlyTTrpAlaGlnAspThrIleLysGluSerGlyThrGluA 584  
|||||  
1720 TGAGCTGTTCGGTGGGCGCAGGACACCATCAAGGAATCCGGTACTGAAG 1769  
|||||  
584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgAsp 600  
|||||  
1770 ATATCTTCAACAAGCAGCGCTGTGCTGGATATGCTGAACGAGACCGCGAT 1819  
|||||  
601 GlyValSerAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa 617  
|||||  
1820 GCGGTGTGAGATCATTCCTCGTCACTGTGGACTGTCTGTCTATTTATGGT 1869  
|||||  
617 lTrpHisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634  
|||||

1870 GTGGCAGCGCATTTTGTGGAAACCGCATTTGATCCACAGATTGAGGACC 1919  
634 rgSerTyrProValGluLeu 640  
|||||  
1920 GCTCCTACCCAGTCGAGCTT 1939  
seq\_name: /STD1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAT61016  
seq\_documentation\_block:  
ID AAT61016 standard; DNA; 30001 BP.  
XX  
AC AAT61016;  
XX  
XX 21-APR-1997 (first entry)  
XX  
Total DNA sequence from cosmid clones LP(2)127 and LP(2)128.  
XX  
Cosmid clone; LP(2)127; LP(2)128; chlortetracycline;  
KW biosynthetic pathway; recombinant; production; antibiotic;  
KW heterologous host; Streptomyces lividans; ss.  
XX  
OS Streptomyces aureofaciens.  
XX  
PN US5589385-A.  
XX  
PD 31-DEC-1996.  
XX  
XX 26-JUL-1990; 90US-0558039.  
XX  
PR 22-SEP-1993; 93US-0125468.  
PR 26-JUL-1990; 90US-0558039.  
PR 26-JUL-1990; 90US-0558040.  
PR 15-JAN-1992; 92US-0821109.  
PR 15-JAN-1992; 92US-0821419.  
XX  
PA (AMCY ) AMERICAN CYANAMID CO.  
XX  
XX Fantini SE, Lotvin JA, Ryan MJ, Strathy N;  
XX WPI; 1997-076853/07.  
XX  
XX DNA encoding tetracycline biosynthetic pathway proteins -  
PT specifically from Streptomyces aureofaciens for expression in  
PT heterologous hosts, specifically S. lividans  
XX  
PS Claim 3; Columns 17-44; 39pp; English.  
XX  
CC The present sequence is the total DNA sequence from cosmid clones  
CC LP(2)127 and LP(2)128, which contains a Streptomyces aureofaciens  
CC (ATCC 13899) DNA that encodes the proteins of the entire  
CC chlortetracycline biosynthetic pathway. The biosynthetic gene,  
CC which can be expressed in heterologous hosts, especially  
CC S. lividans, may be useful in the production of antibiotics.  
XX  
SQ Sequence 30001 BP; 4055 A; 10939 C; 11019 G; 3988 T; 0 other;  
  
alignment\_scores:  
Quality: 633.00 Length: 622  
Ratio: 1.829 Caps: 20  
Percent Similarity: 55.627 Percent Identity: 29.582  
  
alignment\_block:  
US-09-786-474-2 x AAT61016/rev ..  
  
Align seg 1/1 to reverse of: AAT61016 from: 1 to: 30001  
  
1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17  
|||||  
22914 ATGTGCGGAATCGTGGGGTGTGCTGACTACGACGCCCGCGCAACACG 22865  
  
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34  
|||||









```

||||| ||| ... ..
40 ATGTGGGTCTCGGGTGTGGATTACACGGCAAGCTCGACGCA 89
17 evalProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
90 ATTTCCGGCGATCTTCGCATGACCGATACGCTTCGCCCTTGGCGGGCGG 139
34 spAspAlaGlyThrTrpHisAspAlaAspAlaPheGlyPheAsnArg 50
11 ||| ||| ... ..
140 ATGGCGAGGATCTGGAAGCACCGCAACGCCCTGCTGGGTACCGGGCG 189
51 LeuSerIleAlaSpIleAlaHisSerHisGlnProLeuArgTrpGly.. 66
||||| ||| ... ..
190 CTGGCGGTATCGACCTCAGCGGGCGGTGCAGCGGATGCTCTATCGCTT 239
67 .ProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIle 83
||||| ||| ... ..
240 TCCACCGCGCC...GAGGTACACCTCTGTACACCGCGAGGTGT 283
83 yrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPhe 99
||||| ||| ... ..
284 ACAACCATGATCCCTGCGGAGCGGTTCGCCGGCGGACATGATGTTTC 333
100 AsnThrSerGlyAspGlyGluProIleValValGlyPheHisTrpG1 116
||||| ||| ... ..
334 CGCACCGCAGCATACCGAGGTGTCTCGACGCCATATCTGCAATGGGG 383
116 yGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpA 133
||||| ||| ... ..
384 CGAGCGTTGTTGCGAGTACCTCAGCGGGATGTTCCGCTTCGCCGCTTC 433
133 spThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLys 149
||| ... ..
434 ATGGCGCGGCGGCACCTGCTGTGTGTGGCGGACCGCTGGGCATCAAG 483
150 ProLeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGly 166
||||| ||| ... ..
484 CGGCTGTATTACGCGCGCACCGCGAGGAGTCTGTGTGGCTCGGAGAT 533
166 slyThrIleLeu.....GluMetAlaGluMetAsnLeuAspL 180
534 CAAGTCCATCTTGGCGCATCCGGATTCGCCGCCAGCTCGCGCG...G 580
180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
||||| ||| ... ..
581 TCGGCGCTG.....GTCGACCTCTCTGACGCTG 606
197 ProGlu.....ProAspThrLeuHisAlaGlnIleSerArgLeuGluse 211
||| ... ..
607 TCCCGGGGCACCTTCGCACAGCGCGTTCGCCGAGGTCCAGGAACCTGTC 656
211 rGlyCysThrAlaThrValArgProGlyGlyLysLeuGluGlnLysArgT 228
||| ... ..
657 CGGCCACCTGTCTGCTGGCTGCCAATTCACGCGGAAGTTGCGCGCT 706
228 yrPheLysProGlnPheProValGlnLysValValLysGlyLysGlu... 243
||||| ||| ... ..
707 ACTGGGAG.....GTCGCGCGCGGAGGAGCAT 732
244 .....GlnAspLeuPheAspIleAlaGlnValLeuGluAs 256
||||| ||| ... ..
733 GCCGACGACCTGCACAGCACCGTGCAGCGCACCGCGAATGTTCAACCG 782
256 pSerValGluLysHisMetArgAlaAspValThrValGlySerPheLeu 273
||||| ||| ... ..
783 CGCCCTGGGGCGCAATTGCACGCGAGGTTCCGGTGTGTTGCTGCTAT 832
273 erGlyIleAspSerThrAlaIleAlaProLeuAlaLysArg..... 287
||||| ||| ... ..
833 CGGGTGGGCTCGATTTCAGCCGCCCTGACCGGCATCGCCGAGGCATCGCG 882
288 .....HisAsnProAspLeuLeuThrPheThrThrGlyPhe..... 299
||||| ||| ... ..

```

```

883 AAGCGGAGCACGGCGCGACATCAATTCTTCTCGGTGGACTTCGTCGG 932
300 .GluArgGluGlyTyrSerGluValAspVal..... 309
... ||| ... ..
933 CGAGCGCGAGCATTTCCGCGAGCGAGCTCGTCCCGACCGAGGACCG 982
310 .....AlaAlaGluSerAlaAlaAlaIleGlyAlaGluHisIleValLys 324
||| ||| ... ..
983 CGTTCGCCCTGCTGGCGCGAGTACATCGGAGCGCTCATCGCACCGTG 1032
325 IleValSerProGluGluTyr..... 331
||||| ||| ... ..
1033 CTCATCGCAATGCGCAACTGCTCTCGCAACGAGCGCGCAAGAGGTATT 1082
332 .....AlaAsnAlaIleProLysIleMetTrpTyrLeuAspProValA 347
||||| ||| ... ..
1083 CGGGCCCAAGGACGTACCTTTCACCTTCGGCGACATGGAT..... 1122
347 laAspProSerLeuValProLeuTyrPheValAlaAlaGluAlaArgLys 363
... ||| ... ..
1123 .....ACCTCGCTGCACCTGATGTTCCGGGAGATCCGCGG 1158
364 HisValLysValValLeuSerGlyGluGlyAlaAspGluLeuPheGlyC1 380
||||| ||| ... ..
1159 CATTCACGGTGGCATCTCCGCTGAAGCGCGGAGCTGTTCGGTGG 1208
380 yTyrThrIleTyrLysGluProLeuSerLeuAlaProPheGluLysIleP 397
||||| ||| ... ..
1209 CTACGGCTGGTTCGCGCATCCGACGGGTGGCTGCG...GCGCGCTTCC 1255
397 ro.....SerProLeuArgLysGlyLeuGlyLysLeuSerLysValLeu 411
||| ||| ... ..
1256 CCTGGGCTCCAGGTGCGC.....CTG 1278
412 ProAspGlyMetLysGlyLysSerLeuLeuGluArgGlySerMetThrMe 428
... ||| ... ..
1279 CGGCGCGG.....TTCATCGACGCGGTTTC..... 1305
428 tGluGluArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetG 445
||||| ||| ... ..
1306 .....ACCGCGCTGCGATCTCTCTCAGTACCAGC 1336
445 InArgValIleProTrpAlaLysArgGluTrpAspHisArgGluValThr 461
||| ... ..
1337 AGCCGAGCTACGACCATGGCTGCGCCAGGTGCAACACCTGGCGCGGAC 1386
462 AlaProIleTyrAlaGlnSerArgAsnPheAspProValAlaArgMetG1 478
||||| ||| ... ..
1387 AGCCCGGAGGAGCGGGATGCGGAGTTC.....AG 1418
478 nHisLeuAspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaA 495
||||| ||| ... ..
1419 CCACCTGCATCTGAAGCGCTGGATG...GTGCTGTCTCGAACGCAAGG 1465
495 sPLysIleAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAsp 511
||||| ||| ... ..
1466 ATCGCTGAGCATGTGCAACGCGCTGGAGTGGGCTGCCCTACACCGAC 1515
512 LysGluValPheLysValAlaGluThrIleProTyrAspLeuLysIleAl 528
||||| ||| ... ..
1516 CATGAGCTGGTGGATGACGTCTACAACGTGCCCTGTCGATCAAGAGCG 1565
528 aAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGlnIleValP 545
... ||| ... ..
1566 GGACGCGGAGGAGAAGTGGCTCAAGCGGCGCTCGCGGCGCTATGTC 1615
545 roProHisValLeuHisArgLysLysLeuGlyPheProValProMetArg 561
||| ||| ... ..
1616 CGGAAGCGGTGCTCAAGCGCGCCAGAGCCCTTATCCGACTTCTGCCAAC 1665
562 HisTrpLeuAlaGlyAspGluLeuPheGlyTrpAlaGlnAspThrIle 578
||||| ||| ... ..
1666 .....CTCGGCTACGAGGTTTCTCTGCGCGGAGCGTGGCGGCTGCT 1709

```



```

|||||
208787 AAGGAAATTTCTGATCCCTCAAGGGTTCAAGGATCTTTTCCTCAGG..... 208743
251 aglnValLeuGluaspSerValGluLysHisMetArgAlaaspValThrV 268
|||||
208742 .....GCTGTGAAGATAGACATCTTGTGCCGATGTCGCG 208709
268 alGlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProLeu 284
|||||
208708 TGGGTTCTTGCTCAGCGGTGGCCTTGATAGCTCGAGCATAGTC...TGC 208662
285 AlalysArgHis.....AsnProAspLeuLeuThrPheThrG1 298
|||||
208661 GCCATGAGACACCTCCTCCCGATGGTGCAGATAAAGGTGTCTCGCTAGT 208612
298 ypeGluArgGluGlyTyrSerGluValaspValAlaAlaGluSerAlaA 315
|||||
208611 TTTTCCCGGCAAG.....GAAATTTGATGAAGCAAAATATCAG.... 208575
315 laAlaIleGlyAlaGluHisIleValLysIleValSerProGluGluTyr 331
208574 .....AGAAATGTGCTAGACGAA..... 208557
332 AlaAsnAlaIleProLysIleMetTrpTyr..... 341
208556 .....TGCAGGGTAAGCTGGTACAGGACAACCTTTTAAAGTCTGA 208519
342 .....LeuaspAspProv 346
208518 GGATGTTTTGGCTGATATCATATAGTAATCTCATAGAACTCAGGAGGCGCT 208469
346 alAlaAspProSerLeuValProLeuTyrPheValAlaAlaGluAlaArg 362
208468 TTTGACCCCTGAGTATATACGGTCAAGTACAGGGTAATGAAGCTCGCAAGA 208419
363 LysHis...ValLysValValLeuSerGlyGluGlyAlaAspGluLeuPh 378
|||||
208418 GAAACGGAATGAAAGTTCTCTCTCAGCGCCAGCGGAGGATGAGATACT 208369
378 eGlyGlyTyrThrIleTyr..... 384
208368 TGCCGGATACCACTACTTCTTGGCTATTACTTTGTCGAGCTCCTTAGGA 208319
385 .....LysGluProLeuSerLeuAlaProPheGluLysIle 396
208318 AGTTTAAATGGGCAAACTTTGACGGAAATCTTGCCTACAGCGGATC 208269
397 ProSerProLeuArgLysGlyLeuLysLysSerLysValLeuProAs 413
208268 CACGGATCCCTTGTCCTCTCAAGAACATGGTTCTCTACATCTTCCGCT 208219
413 pGlyMetLysGlyLysSerLeuLeuLysGlySerMetThrMetGluG 430
208218 ATGGGTCACCC...AAGAGACTGTGGAGAGGAGATTTCCGTATCTCCGTG 208172
430 luArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetGlnArg 446
208171 AAGAGTTTACAGAGAGTTTAA.....AAGGACCTTCCAAA 208134
447 ValIleProTrpAlaLysArgGluTrpAspHisArgGluValThrAlaPr 463
208133 GAACGTGATGAAAAATAAACTCAACGAGGCTCTCTGTAGCAGA 208084
463 oileTyrAlaGlnSerArgAsnPheAspProValAlaArgMetGlnHisL 480
208083 GACTTACTATTCT.....CTACCTCATC 208061
480 euAspLeuPheThrTrpMetArgLysIleLeuValLysAlaAspLys 496
208060 TCCTCAGGTTT.....GAGCACAAA 208041
497 IleAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysG1 513
|||||

```

```

208040 AACGCAATSCGGTGGAGCATAGAGACAGAGTTCCTTCTCGCACCACGA 207991
513 uValPheLysValAlaGluThrIleProTyrAspLeuLysIleAlaAsnG 530
|||||
207990 GCTCGTTGATACGTTCTCTCTTCCCCCGAATCCAAAGGTTTCGCGCG 207941
530 lyThrThrLysTyrAlaLeuArgArgAlaLeuGluGlnIleValProPro 546
|||||
207940 GTATTACAAAAGTTCTACTCGGTAAAGGTCTTAAGGGCATCTCGCGGAT 207891
547 HisValLeuHisArg...LysLysLeuGlyPheProValProMetArgHi 562
|||||
207890 GAAATACGGAATAGGCTAGCAAGGTAGGTTTGCAGCCCTGCACAAAGA 207841
562 strPheAlaGlyAspGlu.....LeuPheGlyTrpAlaGlnAspThrI 577
207840 CATCCTTAAGACTAATGAAGGCGACAGTTTGGTGGCTGTGTATAGACT 207791
577 leLysGluSerGlyThrGluAspIlePheAsnLysGlnAlaValLeuAsp 593
|||||
207790 CTGAAAGTTTCAAAAACGGCCTTACTGGGATTATCGAAAAGTATTAAAG 207741
594 MetLeuAsnGluHisArgAspGlyValSerAspHisSerArgArgLeuTr 610
|||||
207740 ATGTTTGAGGAGCACGTGTCTGGCAAGAAAACTGGAGCCAAAGAGCTGTG 207691
610 pThrValLeuSerPheMetValTrpHisGlyIlePheValGluAsn 625
207690 GAAGTTTATATCATCGAGCTGTGGCTCAGGAGTGGATGGATAAT 207645
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA58471
seq_documentation_block:
ID AAA58471 standard; DNA; 58857 BP.
XX
AC AAA58471;
XX
DT 31-OCT-2000 (first entry)
XX
DE Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS Streptomyces verticillius.
XX
FH Key Location/Qualifiers
CDS 223..564
FT /*tag= a
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 30; encodes AAB07556"
FT CDS 561..2309
FT /*tag= b
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 29; encodes AAB07557"
FT CDS 2767..3486
FT /*tag= c
FT /note= "ORF 28; encodes AAB07558"
FT CDS 3527..5593
FT /*tag= d
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 27; encodes AAB07559"
FT CDS 5806..12294
FT /*tag= e
FT /note= "ORF 26; encodes AAB07560"
FT CDS 12291..15491
FT /*tag= f
FT /note= "ORF 25; encodes AAB07561"
FT CDS 15488..21013
FT /*tag= g
FT /note= "ORF 24; encodes AAB07562"
FT CDS 21010..24666

```

```

FT FT      /*tag= h
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 23; encodes AAB07563"
FT FT      24663..32690
FT FT      /*tag= i
FT FT      /note= "ORF 22; encodes AAB07564"
FT FT      34893..34830
FT FT      /*tag= j
FT FT      /note= "ORF 21; encodes AAB07565"
FT FT      34827..35804
FT FT      /*tag= k
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 20; encodes AAB07566"
FT FT      35818..37302
FT FT      /*tag= l
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 19; encodes AAB07567"
FT FT      37299..39215
FT FT      /*tag= m
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 18; encodes AAB07568"
FT FT      39301..47181
FT FT      /*tag= n
FT FT      /note= "ORF 17; encodes AAB07569"
FT FT      47178..49985
FT FT      /*tag= o
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 16; encodes AAB07570"
FT FT      49982..51001
FT FT      /*tag= p
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 15; encodes AAB07571"
FT FT      50998..52386
FT FT      /*tag= q
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 14; encodes AAB07572"
FT FT      52383..52946
FT FT      /*tag= r
FT FT      /note= "ORF 13; encodes AAB07573"
FT FT      53018..54190
FT FT      /*tag= s
FT FT      /note= "ORF 12; encodes AAB07574"
FT FT      54187..55824
FT FT      /*tag= t
FT FT      /note= "ORF 11; encodes AAB07575"
FT FT      55821..56093
FT FT      /*tag= u
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 10; encodes AAB07576"
FT FT      56090..57586
FT FT      /*tag= v
FT FT      /transl_except= (pos: 1..3, aa: Met)
FT FT      /note= "ORF 9; encodes AAB07577"
FT FT      57583..58857
FT FT      /*tag= w
FT FT      /note= "ORF 8; encodes AAB07578"
FT FT
FT FT      WO200040704-A1.
FT FT      13-JUL-2000.
FT FT
FT FT      06-JAN-2000; 2000WO-US00445.
FT FT
FT FT      06-JAN-1999; 99US-0115435.
FT FT      05-FEB-1999; 99US-0118848.
FT FT      05-JAN-2000; 2000US-0477962.
FT FT
FT FT      (REGC ) UNIV CALIFORNIA.
FT FT
FT FT      Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
FT FT      WPI; 2000-465974/40.
FT FT      P-FSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,

```

```

DR AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR AAB07576, AAB07577, AAB07578.
XX
XX New bleomycin gene cluster components useful for peptide and/or
PT polyketide metabolites, especially bleomycin, production and for
PT chemically modifying biological molecules -
XX
XX Claim 8; Page 97-136; 162pp; English.
XX
XX The present sequence represents the BLM (Bleomycin) gene cluster,
CC containing open reading frames (ORFs) 8-30. The proteins encoded
CC by the gene cluster are useful for producing peptides and/or polyketide
CC metabolites, especially bleomycin or bleomycin analogues. They are
CC also useful for chemically modifying biological molecules to produce
CC branched methyl groups, and for coupling amino acids and fatty
CC acids. They may be reacted with an apo-carrier protein and coenzyme A
CC to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC domains can be used individually or collectively to produce
CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC microbial metabolites. The BLM gene cluster may also be used to produce
CC sugars.
XX
XX Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;
SQ

```

alignment\_scores:

|                     |        |                   |        |
|---------------------|--------|-------------------|--------|
| Quality:            | 516.00 | Length:           | 642    |
| Ratio:              | 1.487  | Gaps:             | 23     |
| Percent Similarity: | 54.050 | Percent Identity: | 27.726 |

alignment\_block:

```

US-09-786-474-2 x AA58471 ..
Align seg 1/1 to: AAA58471 from: 1 to: 58857
1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
37308 GTGTGGCGCATCTGGCGATCCGCTCCGCGCGGCGGACTCGAGGCG... 37355
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
|||:::|||||:::|||||:::|||||:::|||||
37356 ....GTTGAACCTACCGCGCGCGATGGCGGACCTGCGCGCGCGCGCGCG 37401
34 sAspAlaGlyThrTrp.....HisAspAlaAspAlaAlaPheGlyPhe 48
||||||||:::|||||:::|||||:::|||||:::|||||
37402 ACGCGGAAGCACCTCGGTCTCGCCCGCGCGCGCGCGCGCGCGCGCGCAC 37451
49 AsnArgLeuSerIleLeuAspIleAlaHisSerHisGlnProLeuArgTr 65
:::|||||:::|||||:::|||||:::|||||:::|||||
37452 ACCCGGCTCGCGTGTATCGCCCGCGGACGCGGCGCGCGCGGTC.... 37496
65 pGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluI 82
|||:::|||||:::|||||:::|||||:::|||||
37497 .....GCCGCGCGCGGACGCGCACCGCTCCGCGCTCGTGTACAGCGCGAGT 37539
82 LeTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThr 98
:::|||||:::|||||:::|||||:::|||||:::|||||
37540 TCTACGGCTACCGGAGATCCGCGCGGAATCGCGCGCGCGCGCGCGCGCGG 37589
99 PheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisIleTr 115
|||:::|||||:::|||||:::|||||:::|||||:::|||||
37590 TTCCGCACCGCGGACGACAGCGAG.....ATCGCGCTCCACCTGTA 37630
115 p.....GlyGluSerValValGluHisLeuArgGlyMetPheGlyI 129
:::|||||:::|||||:::|||||:::|||||:::|||||
37631 CCTGCGGAGACGCGCGCGGCGACTGGAGCGGCTGCGCGCGCGAGTTCCGCT 37680
129 leAlaIleTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGln 145
:::|||||:::|||||:::|||||:::|||||:::|||||
37681 TCGTCTCTCTGGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 37730
146 PheGlyIleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPh 162

```

[illegible]

```

3583  CGCACCTGGCCGTCACCCAGCCCGTCGCCGCCCTGCTCCGC..... 38622
434  yAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrpA 451
38625  ....|||CCCGACT 38631
451  laLysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGln 467
|||:::|||||::|
38632  TCGCCGCGCAACTGGCCGCGCGACGCGCGCGCCCTGCTCGCGCC 38681
468  SerArgAsnPhe.....AspProValAlaArgMetClnHisLe 480
:::|||||:::
38682  GCGCGCGGCTGCTCGCGCGCGCGCCCGCGCACACCGCCACCTACCT 38731
480  uAspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaAspLysI 497
|||||::| |||::|
38732  CTTCCGAAGACCTGGCTGCCCGC...TACCTGCTCGCGCGCGACGCC 38778
497  leAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysGlu 513
:::|||||:::|||||:::|||||:::
38779  TCGACGCGCGCCAGCGCGCTCGAGGTGCGGCTCCCTTCGCACCAACC 38828
514  ValPheLysValAlaGluThrIlePro.....TyrAspLeuLysII 527
:::|||||:::|||||
38829  CTCTTCGACCTGCTCCGCGCACACCCCGCGCGCTGGTAGCAC..... 38870
527  eAlaAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGlnIleV 544
:::||||| ||||| ||||| |||||:::
38871  .AGGACGGCACCGCAAGTACCCGCTCGGGCGCGCATCGGCCACCGGC 38919
544  alProProHisValLeuHisArgLysLysLeuGlyPheProValPromet 560
:::|||||::: ||||| |||||
38920  TCGCGCGGAGGTGACCGAGGCGCCAAACAGGGCTTCCTCGCACCTCCG 38969
561  ArgHisTrpLeuAlaGlyAspGluLeuPheGlyTrpAlaGlnAspThrII 577
:::||||| |||::: |||:::
38970  .....ATGGCGGAGGACGACACCCCTC.....CTCGACGCCCT 39001
577  eLysGlu.....SerGlyThrGluAspIlePheAsnL 588
:::|||||::: |||::: |||:::
39002  GCGGACGCGCTCGCGGACCGGGCGCGGCGACGACCCCTTCTTCGACC 39051
588  ysGlnAlaValLeuAspMetLeuAsn 596
:::|||||::: |||:::
39052  CGCACGCGCTCGCGCGCCCTGCTGGAC 39077

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: AAC40911
seq_documentation_block:
ID AAC40911 standard; DNA; 2164 BP.
XX AC AAC40911;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 29975.
XX KW Hybridisation assay; genetic mapping; gene expression control
KW protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.

```

PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132866.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.



```

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 468.00      Length: 617
  Ratio: 1.436        Gaps: 27
  Percent Similarity: 52.836   Percent Identity: 26.580

alignment_block:
US-09-786-474-2 x AAC40911 ..

Align seg 1/1 to: AAC40911 from: 1 to: 2164

```

```

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla... 16
|||||...|||||...|||||...|||||...|||||...
196 ARGTTGGAATACTGGCGTGTAGGATGTTCCGATGATTCAGGCCAA 245
17 .....PheValProAlaLeuGluArgAlaLeuProCysMetArgHisa 31
|||||...|||||...|||||...|||||...
246 GAGATTTCGTGTTCTCGAGCTTCTCGCAGATTG.....AGGCACA 286
31 rgGlyProAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
|||||...|||||...|||||...|||||...|||||...
287 GAGGACCTGACTGGAGTGGATTATATCAGACGGAGATAATTACTGGCC 336
48 PheAsnArgLeuSerIleleaspIleAlaHisSerHisGlnProLeuAr 64
|||||...|||||...|||||...|||||...|||||...
337 CATCAACGCTCTGCGCTCATCGCTCCTCGGTGATCAACCTCTT... 384
64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
|||||...|||||...|||||...|||||...
385 .....TTCAACGAGGACAGACCATTTGTCACGGTGAACGGAG 424
81 luIleTyrAsnTyrValGluLeuArgLysGlyLeuSerAspLeuGlyTyr 97
|||||...|||||...|||||...|||||...|||||...
425 AGATTATAACCATCAGAGCTGAGAAACGCTCAAGAAAT.....CAC 468
98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisH 114
|||||...|||||...|||||...|||||...|||||...
469 AAGTTCCTGCTAGTGGTGTGATGTTGAAGTCAATGCTCACTTGTACGAGGA 518
114 sTrpGlyGluSerValGluHisLeuArgGlyMetPheGlyIleAlaI 131
|||||...|||||...|||||...|||||...|||||...
519 GPATGGTGGATTTGTTGATATGTTGGATGGAATCTTCTCTTGTGT 568

```

```

131 lTrpAspThrLysGlyLysSerLeuPheLeuAlaArgAspGlnPheGly 147
:::|||||...|||||...|||||...|||||...|||||...
569 TGCTCGACACAGAGATAACTCTTCATGCTGGTGGTGGTATGCCGATGGT 618
148 lIleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
:::|||||...:::|||||...:::|||||...:::|||||...
619 GTCACTTCGCTCTACATTGGTGGGACTAGACGGATCTGTGTGGATATC 668
163 rSerGluLysLysThrIleLeuGluMetAlaGluMetAsnLeuAspL 180
|||||...|||||...
669 TTCAGAGATGAAA..... 681
180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
|||||...:::|||||...
682 ..GGCTTAACGAT...GATTGTGAGCATTC..... 708
197 ProGluProAspThrLeu.....HisAlaGlnIleSerArgLeuGl 210
:::|||||...:::|||||...
709 .....GAAACGTTTCTCCAGGTCATTTTATTCAAGCAAGTTA... 747
210 uSerGlyCysThrAlaThrValArgProGlyGlyLysLeuGluGlnLysA 227
|||||...|||||...
748 .....GGAGGG.....TTTAAGC 760
227 rgTyrPheLysProGlnPheProValGlnLysValValLysGlyLysGlu 243
:::|||||...:::|||||...
761 AATGTTATAATCTCTCTGGTTCATGAATCTGTTCTTCAACGCCTTAT 810
244 GlnAspLeuPheAspArgIleAlaGlnValLeuGluAspSerValGluLy 260
:::|||||...:::|||||...
811 GAGCCTCTT.....GCATAGACGCGCCTTTGAAAACGCTGTGATTAA 854
260 sHisMetArgAlaAspValThrValGlySerPheLeuSerGlyGlyIleA 277
|||||...:::|||||...
855 GCGGTGATGACTGATTCCTTCATTTGGAGTTTGTCTCTCTGTTGCTGTG 904
277 spSerThrAlaIleAlaProLeuAlaLysArgHis..... 288
|||||...:::|||||...
905 ATTCCTTCCTTGTTCCTTCATCCTGTCAGCTGCTGCGCGGTACTAAG 954
289 .....AsnProAspLeuLeuThrPheThrThrGlyPheGl 300
:::|||||...:::|||||...
955 GCGGTAAGCAATGGGTGCTCAGCTCCTCTTTCGTTGGTGGTCTT... 1002
300 uArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaAlaI 317
|||||...:::|||||...
1003 .....GAGGCTCACCGGACTTGAAGCAGGGAAGAGGTGGCGGAATAT 1048
317 leGlyAlaGluHisIleValLysIleValSerProGluLutyrAlaAsn 333
:::|||||...:::|||||...
1049 TGGGACGCTGCACACGAGTTCACCTTCTCGTGCAGGACGGATGAT 1098
334 AlaIleProLysIleMetTrpTyrLeuAsp.....AspProValAlaAs 348
|||||...:::|||||...
1099 GCGATTGAGGATGATTTACCATGTTGAGACCTATGATGTGACGACTAT 1148
348 pProSerLeuValProLeuTyrPheValAlaAlaGluAlaArgLys...H 364
|||||...:::|||||...
1149 CAGACGACACACCGATGTTCTTGTATGTCGCCGAAATCAAGTCTCTAG 1198
364 lsValLysValValLeuSerGlyGlyAlaAspGluLeuPheGlyGly 380
|||||...|||||...
1199 GGTCAAGATGTTCTCTCCGCGAAGGTGCGGACGAGATCTTTGGAGGG 1248
381 TyrThrIleTyrLysGluProLeuSerLeuAlaProPheGluLysIlePr 397
|||...:::|||||...
1249 TAC...CTCTAT.....TTCCACAAGGCACC 1271
397 oSerProLeuArgLysGlyLeuGlyLysLeuSerLysValLeuProAspG 414
|||||...
1272 TAAC..... 1275
414 lyMetLysGlyLysSerLeuLeuGluArgGlySerMetThrMetGluGlu 430

```

```
1275 ..... 1275
431 ArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgVa 447
1275 ..... 1275
447 IleProTrrAlaLysArgGluTrpAspHisArgGluValThrAlaProI 464
1276 ..... 1309
464 letYrAlaGlnSerArgAsnPheAspProValAlaArgMetGlnHisLeu 480
1310 TCAAGGCTCTTCAAG...TATGAC..... 1332
481 AspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaAspLysI 497
1333 .....TGCTTAAGAGCCCAACAATC 1352
497 easnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysGluV 514
1353 TACCTCTGCGTTGGACTAGAGCAGCTGTTCCTTCCCTTGACAAGACT 1402
514 alPheLysValAlaGluThrIleProTyrAspLeuLysIle..... 527
1403 TCATCAACACAGCTATGCTCTCGACCCCTGAATCCAAGATGATCAAGCCA 1452
528 ...AlaAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGln... 542
1453 GAGGAAGAGGAGTACGAGAATGGTCTTAAGAGAGCCCTTGACGACGA 1502
543 .....IleValProProHisValLeuHisArgLysLysLeuGlyP 556
1503 AGAACGTCCTTATCTACCAAAACACATTCTCTACACACAGAAAGACAGT 1552
556 heProValProMetArgHisTrpLeuAlaGlyAspGluLeu...PheGly 571
1553 TC.....AGTCATGGTGTGGCTACAGT 1575
572 TrpAlaGlnAspThrIleLysGluSerGlyThrGluAspIlePheAsnLys 588
1576 TGGATC...GATGGCCTGAAGATCVCGGCTGCTCAAAATGTCAANTGACAA 1622
588 s 588
1623 G 1623
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA50959
seq_documentation_block:
AA50959 standard; DNA; 2162 BP.
AC
AAC50959;
XX
XX
DT 18-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66760.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS
OS Arabidopsis thaliana.
XX
XX
PN EP1033405-A2.
PD
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
PF 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
```

```
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
```

|    |              |                |
|----|--------------|----------------|
| PR | 16-JUL-1999; | 99US-01144085; |
| PR | 16-JUL-1999; | 99US-01144086; |
| PR | 16-JUL-1999; | 99US-01144325; |
| PR | 19-JUL-1999; | 99US-01144331; |
| PR | 19-JUL-1999; | 99US-01144332; |
| PR | 19-JUL-1999; | 99US-01144333; |
| PR | 19-JUL-1999; | 99US-01144334; |
| PR | 19-JUL-1999; | 99US-01144335; |
| PR | 20-JUL-1999; | 99US-01144352; |
| PR | 20-JUL-1999; | 99US-01144632; |
| PR | 20-JUL-1999; | 99US-01144884; |
| PR | 21-JUL-1999; | 99US-01148114; |
| PR | 21-JUL-1999; | 99US-01145086; |
| PR | 21-JUL-1999; | 99US-01145088; |
| PR | 22-JUL-1999; | 99US-01145085; |
| PR | 22-JUL-1999; | 99US-01145087; |
| PR | 22-JUL-1999; | 99US-01145089; |
| PR | 22-JUL-1999; | 99US-01145192; |
| PR | 23-JUL-1999; | 99US-01145345; |
| PR | 23-JUL-1999; | 99US-01145218; |
| PR | 23-JUL-1999; | 99US-01145224; |
| PR | 26-JUL-1999; | 99US-01145276; |
| PR | 27-JUL-1999; | 99US-01145913; |
| PR | 27-JUL-1999; | 99US-01145918; |
| PR | 27-JUL-1999; | 99US-01145919; |
| PR | 28-JUL-1999; | 99US-01145951; |
| PR | 02-AUG-1999; | 99US-01146386; |
| PR | 02-AUG-1999; | 99US-01146388; |
| PR | 02-AUG-1999; | 99US-01146389; |
| PR | 02-AUG-1999; | 99US-01147038; |
| PR | 04-AUG-1999; | 99US-01147204; |
| PR | 04-AUG-1999; | 99US-01147302; |
| PR | 05-AUG-1999; | 99US-01147192; |
| PR | 05-AUG-1999; | 99US-01147260; |
| PR | 08-AUG-1999; | 99US-01147303; |
| PR | 08-AUG-1999; | 99US-01147416; |
| PR | 08-AUG-1999; | 99US-01147493; |
| PR | 13-AUG-1999; | 99US-01148684; |
| PR | 16-AUG-1999; | 99US-01149368; |
| PR | 17-AUG-1999; | 99US-01149175; |
| PR | 18-AUG-1999; | 99US-01149426; |
| PR | 20-AUG-1999; | 99US-01149722; |
| PR | 20-AUG-1999; | 99US-01149723; |
| PR | 20-AUG-1999; | 99US-01149929; |
| PR | 23-AUG-1999; | 99US-01149902; |
| PR | 23-AUG-1999; | 99US-01149930; |
| PR | 25-AUG-1999; | 99US-01150566; |
| PR | 26-AUG-1999; | 99US-01150884; |
| PR | 27-AUG-1999; | 99US-01151065; |
| PR | 27-AUG-1999; | 99US-01151066; |
| PR | 27-AUG-1999; | 99US-01151080; |
| PR | 30-AUG-1999; | 99US-01151303; |
| PR | 31-AUG-1999; | 99US-01151438; |
| PR | 01-SEP-1999; | 99US-01151930; |
| PR | 10-SEP-1999; | 99US-01152363; |
| PR | 07-SEP-1999; | 99US-01153070; |
| PR | 13-SEP-1999; | 99US-01153758; |
| PR | 15-SEP-1999; | 99US-01153758; |
| PR | 16-SEP-1999; | 99US-01154039; |
| PR | 20-SEP-1999; | 99US-01154717; |
| PR | 22-SEP-1999; | 99US-01155139; |
| PR | 23-SEP-1999; | 99US-01155486; |
| PR | 24-SEP-1999; | 99US-01155659; |
| PR | 28-SEP-1999; | 99US-01156458; |
| PR | 29-SEP-1999; | 99US-01156596; |
| PR | 04-OCT-1999; | 99US-01157117; |
| PR | 05-OCT-1999; | 99US-01157753; |
| PR | 06-OCT-1999; | 99US-01157865; |

```

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 467.00      Length: 617
    Ratio: 1.437         Gaps: 27
    Percent Similarity: 52.674      Percent Identity: 26.580

alignment_block:
US-09-786-474-2 x AAC50959 ..

Align seg 1/1 to: AAC50959 from: 1 to: 2162

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla... 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 ATGTGTGGAATACTTCGCGTGTAGGATGTTCCGATGATTCTCAGCGCAA 244

17 .....PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
||| |||||:||||| ||| |||||:|||||
245 GAGAGTTCGTGTTCTTCGAGCTTTCGCGAGATTG.....AGGCACA 285

31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaPheGly 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 GAGACCTGACTGAGTGCGCTTATATCATCAGACGGAGATAATTACTTGCC 335

48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 CATCAACAGCTTCGCGTCATCGATCCTCTCCGCTGATCAACCTCTT... 383

64 gTTPGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
|||||:|||||:|||||:|||||:|||||:|||||:
384 .....TTCAACGAGGACAAGACCATGTTGTCACGGTGAACGGAG 423

81 luileTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 AGATTTATACCATGAGGAGCTGAGAAACAGCTCTGAAGANT.....CAC 467

98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisH 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
468 AAGTTCGCTACTGCTAGTGAATGTTGAAGTCATTGCTCACTTGTACCGAG 517

114 sTTPGlyCyluSerValValGluHisLeuArgGlyMetPheGlyIleAla 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
518 GTATGCTGTGGATTTTGTGATGATGTTTCGATGGAATATCTCCTTTGGT 567

```

```

131 letrpAspThrLysGluLysSerLeuPheLeuAlaAArgAspGlnPheGly 147
   :: |||||:::|||||:::|||||:::|||||:::|||||
568 TGCTCAGACACAGAGTACTCTTTTCATGGCTCGTATCGCATGGT 617
   :: |||||:::|||||:::|||||:::|||||:::|||||
148 IleLysProLeuPheTyrAlaThrGluHisGlyThrValPhe...Se 163
   :: |||||:::|||||:::|||||:::|||||:::|||||
618 GTCACCTTCGCTACATTTGTTGGGACTAGACGGATCTGTGGATATC 667
   :: |||||:::|||||:::|||||:::|||||:::|||||
163 rSerGluLysLysThrIleLeuGluMetAlaGluMetAsnLeuAspL 180
   ||||| |||
668 TTCAGAGATGAAA..... 680
180 euGluLysAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
   |||||:::|||||:::|||||:::|||||:::|||||
681 ..GGCCTAAACGAT...GACTGTGACATTC..... 707
197 ProGluProAspThrLeu.....HisAlaGlnIleSerArgLeuG1 210
   ::|||:::|||||:::|||||:::|||||:::|||||
708 .....GAAACGTTTCCTCCAGGTCATTTTATTCAAGCAAAATTA.. 746
210 userGlyCysThrAlaThrValArgProGlyLysLysLeuGluGlnLysA 227
   |||||
747 .....GGAGGG.....TTTAAGC 759
227 rGlyrPheLysProGlnPheProValGlnLysValValLysGlyLysGlu 243
   ::|||:::|||||:::|||||:::|||||:::|||||
760 AATGGTATATCCTCCTCGTCAATGAATCTGTTCTTCAACGCCCTTAT 809
244 GlnAspLeuPheAspArgIleAlaGlnValLeuGluAspSerValGluLy 260
   :: |||||:::|||||:::|||||:::|||||:::|||||
810 GAGCCTCTT.....CGCATAGACGCGCTTTGAAAACGCTGTGATTAA 853
260 sHisMetArgAlaAspValThrValGlySerPheLeuSerGlyLysIleA 277
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
854 CGGGTTGATGACTGATGTTCCATTTGGAGTTTGTGCTCTCTGGTGGTCT 903
277 spSerThrAlaIleAlaProLeuAlaLysArgHis..... 288
   |||||:::|||||:::|||||:::|||||:::|||||
904 ATCTTCCCTTGTGCTCCATCCTCAGTCGACGTCACCTTGGCGGTACTAAG 953
289 .....AsnProAspLeuLeuThrPheThrThrGlyPheG1 300
   ::|||:::|||||:::|||||:::|||||:::|||||
954 CGCGCTAAGCAATGGGGTCTCTCAGCTCCATTCCTTTTGGTGGTCT.. 1001
300 uArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaAlaI 317
   |||||:::|||||:::|||||:::|||||:::|||||
1002 ....GAGGCTCACCGACTTGAGCGAGGAAAGAGGTGGCGGAATATT 1047
317 leGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsn 333
   ::|||:::|||||:::|||||:::|||||:::|||||
1048 TGGGAGCGGTGCACACGAGTTCACCTTCTCGGTGCAGGACGGGATTGAT 1097
334 AlaIleProLysIleMetTrpTyrLeuAsp.....AspProValAlaAs 348
   |||||:::|||||:::|||||:::|||||:::|||||
1098 GCGATAGAGATGTGATTTACCATGTTGAGACCTATGATGTGACGACTAT 1147
348 pProSerLeuValProLeuTyrPheValAlaAlaGluAlaArgLys...H 364
   ::|||:::|||||:::|||||:::|||||:::|||||
1148 CAGAGCGACACACCATGTTCTTGTATGTCGCCGAAAATCAAGTCTCTAG 1197
364 isValLysValValLeuSerGlyGluGlyAlaAspGluLeuPheGlyGly 380
   |||||:::|||||:::|||||:::|||||:::|||||
1198 GGGTCAAGATGGTCTCTCCGGCGAAGGTGCGGACGAGATCTTTGGAGGG 1247
381 TyrThrIleTyrLysGluProLeuSerLeuAlaProPheGluLysIlePr 397
   |||||:::|||||:::|||||:::|||||:::|||||
1248 TAC...CTCTAT.....TTCCACAAGGCACC 1270
397 oSerProLeuArgLysGlyLeuGlyLysLeuSerLysValLeuProAspG 414
   |||||
1271 TAAC..... 1274

```

```

414 lyMetLysGlyLysSerLeuLeuGluArgGlySerMetThrMetGluGlu 430
1274 ..... 1274
431 ArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgVa 447
1274 ..... 1274
447 lleProTTPAlaLysArgGluTrpAspHisArgGluValThrAlaProI 464
   |||||:::|||||:::|||||:::|||||:::|||||
1275 .....AGAAAGAGTTT...CACCAAGAAACTTGTCCGAAGA 1308
464 letYrAlaGlnSerArgAsnPheAspProValAlaArgMetGlnHisLeu 480
   || ||| :::: ||||
1309 TCAAGGCTCTTCACAAG...TATGAC..... 1331
481 AspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaAspLysII 497
   ::|||:::|||||:::|||||:::|||||:::|||||
1332 .....TGCTAAGAGCCCAACAATC 1351
497 eAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysGluV 514
   ::|||:::|||||:::|||||:::|||||:::|||||
1352 TACCTCTGCTTTGGACTAGAGCAGCTGTCTTCTTCCCTTGACAAGACT 1401
514 alPheLysValAlaGluThrIleProTyrAspLeuLysIle..... 527
   ::|||:::|||||:::|||||:::|||||:::|||||
1402 TCATCAACACAGCTATGCTCTCGACCCTGAATCCAAGATGATCAAGCCA 1451
528 ...AlaAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGln.. 542
   ::|||:::|||||:::|||||:::|||||:::|||||
1452 GAGGAAGGAAGATCGAGAAATGGGTCTTAAGGAGAGCCTTTGACGACGA 1501
543 .....IleValProHisValLeuHisArgLysLysLeuGlyP 556
   ::|||:::|||||:::|||||:::|||||:::|||||
1502 AGAACCTCTCTTATCTACCAAAACACATCTCTACACACAGAAACACAGT 1551
556 heProValProMetArgHisTrpLeuAlaGlyAspGluLeu...PheGly 571
   || ||| :::: ||||
1552 TC.....AGTGATGGTGTGGCTACAGT 1574
572 TrpAlaGlnAspThrIleLysGluSerGlyThrGluAspIlePheAsnLy 588
   |||| ::|||:::|||||:::|||||:::|||||:::|||||
1575 TGGATC...GATGGCTGAAAGATCACGCTGCTCAAAATGTCAATGACAA 1621
588 s 588
1622 G 1622
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAV81946
seq_documentation_block:
ID AAV81946 standard; DNA; 40138 BP.
XX
AC AAV81946;
DT 19-OCT-1999 (first entry)
XX
XX V. marinus PKS-like cluster comprising ORFs 6,7,8 and 9.
DE
KW Polyketide-like synthesis; PKS; PKS-like gene; PUFA; DHA; transgenic;
KW poly-unsaturated fatty acid; eicosapentenoic acid; docosahexanoic acid;
KW EPA; oil; dietary supplement; infant feeding formulation; malnutrition;
KW intravenous feeding formulation; cooking oil; fat; anti-inflammatory;
KW cholesterol; open reading frame; ORF; ss.
XX
OS Vibrio marinus.
XX
FH Key Location/Qualifiers
FT CDS 17394..25352
FT FT /*tag= a
FT FT /product= "ORF 6"
FT FT 25509..28160
FT FT /*tag= b

```





```

11200 CGGCCCCAGATTGGTCAGGTATCTATACAAGTGAATGCTATTTAGCTT 11151
48 PheAsnArgLeuSerIleAlaAspIleAlaHisSerHisGlnProLeuAr 64
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11150 CATGAACGCTCTTGCTATGTCGATGTAATAATGGTGCACAGCCGCTA... 11103
64 gTPrGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
11102 .....TATAATGAAGAAAAGACACACAGCTACTTGCCTTAATGGTG 11063
81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
11062 AAATTTATAACCAACAAAGATTAAAGAAAACGCTG...AACGTAGACTTT 11016
98 ThrPheAsnThrSerGlyAspGlyProIleValValGlyPheHisH1 114
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
11015 GAATTCACCAACAGATCAGATTGCGAAGTGATCTTGGCTATATAAGA 10966
114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10965 AAAAGGCACACAATTTTAGATGATCTAAATGGTATTTTCTTTCGCTT 10916
131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10915 TATATGACAAACACAGAGCAGCTTACCTGATTGGCCGTGACCATATCG 10866
148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...se 163
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10865 ATCATCCCTCTTTATACGGGTATGATGAACACCGTAACTTTATGTGC 10816
163 rSerGluLysLysThrIleLeuGluMetAlaGluMetAsnLeuAspL 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10815 CTCTGAATGAAGACACTAGTGCCAATTTGTACTCAA..... 10779
180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10778 .....ATTGAAGAATTTCCAGCAGGTCACTTACCTG 10749
197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10748 TGGAGTAAGGAT..... 10737
213 sThrAlaThrValArgProGlyGlyLysLeuGluGlnLysArgTyrPheL 230
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10736 ...GGTGTAGTCAGCCT.....TATTATC 10715
230 ysProGlnPheProValGlnLysValValLysGlyLysGluGlnAspLeu 246
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10714 AA.....CGTGATTGGAAGAT 10698
247 PheAspArgIleAla.....GlnVa 253
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10697 TTTGATAACGTTGCACAAAATGGTGGTGATAAACCGTTGTTAAGCAAG 10648
253 lLeuGluAspSerValGluLysHisMetArgAlaAspValThrValGlys 270
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10647 TTTGGAGATGCAGCTTAAACGTCATTAATGCTGTGATGCTACCTTACGCG 10598
270 erPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProLeuAlaLys 286
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10597 TATTATTATCAGCGGGTTAGATTCTGTTATTTCAGCAATCACCCAA 10548
287 ArgHisAsn..... 289
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10547 CAGTATGCAAAACGTCGTATTGAAGACGGTGGTAAACCGCAAGCATGGTG 10498
290 .ProAspLeuLeuThrPheThrThrGlyPheGluArgGluGlyTyrSerG 306
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10497 GCCACAACCTACACTCTTCTCTGTAGGCCTA.....AATGGTTCTCCAG 10454
306 luValAspValAlaAlaGluSerAlaAlaAlaIleGlyAlaGluHisIle 322
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10453 ATTTAGCGCTGCACAAAAGTGGCAGATCATCTTAGGTACTATCATCAT 10404

```

```

323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe 339
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10403 TCAATTGAATTACTGTGCAAGAAGGTATCGACGCTTTACGCGACGTGAT 10354
339 tTrpTyrLeuAsp.....AspProValAlaAspProSerLeuValProL 354
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10353 TTACCACATTGAACCTACGATGTAAACAATTCGTGCATCAACCCCTA 10304
354 euTyrPheValAlaAlaGluAlaArg....LysHisValLysValValLeu 369
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10303 TGTACCTAATGCGCAGTAAATTAAGCCATGGGATTAAGATGGTGCTT 10254
370 SerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysG 386
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10253 TCTGGTGAAGGTGCTGATGAACTGTTCCGTGTTAC....TTGTAC.... 10212
386 uProLeuSerLeuAlaProPheGluLysIleProSerProLeuArgLysG 403
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10211 .....TTCCACAAAAGCGCA..... 10197
403 lyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGlyLysSer 419
10197 ..... 10197
420 LeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAl 436
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10196 .....ANTGC 10192
436 aArgSerPheAsnPheGluGlnMetGlnArgValIleProTrpAlaLysA 453
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10191 GAAAGATTCATGAAGACGGTACGTAAGTA..... 10158
453 rgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGlnSerArg 469
10158 ..... 10158
470 AsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMe 486
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10157 .....AATAATTACATATGTTTGATGCTT 10132
486 tArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerL 503
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10131 ACGT.....GCAACAAATCTATGCGACGTTGGGCA 10100
503 euGluLeuArgValProPheLeuAspLysGluValPheLysValAlaGlu 519
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10099 TCGAAGCACGTTCTCTTCTTAGATAAAGAATTTGTCATCGCTCAATG 10050
520 ThrIleProTyrAspLeuLysIleAlaAsnGly.....ThrThrLysTy 534
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10049 CGTCTAAACCCCTGAACATAAATGATCAGCGCGACCGCATTTGAGAAAA 10000
534 rAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHisA 551
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
9999 CATCATTCGAGAAGCATTTGAAGATTTACTACCTGAAGAAATTTGATGCG 9950
551 rgLysLysLeuGlyPhe 556
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
9949 GTCAAAAAGAACAAATTC 9933

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAI64984

seq\_documentation\_block:

ID AAI64984 standard; DNA; 41587 BP.

XX AAI64984;

XX 04-DEC-2001 (first entry)

XX Moritella marina icosapentaenoic acid biosynthesis enzyme DNA #1.

XX Icosapentaenoic acid biosynthesis; docosahexanoic acid productivity;







**THIS PAGE BLANK (USPTO)**

OM of: US-09-786-474-2 to: EST:\* out\_format : pfs

Date: Sep 20, 2002 7:47 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US09786474/runat.18092002.132910.5427/app\_query.fasta\_1.706  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -CAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.500  
-DEPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-LIST=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEARSIZE=500  
-INLEN=0 -MAXLEN=2000000000 -USER=US09786474\_@CGNL\_13007  
-ICPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-786-474-2

Query length: 640

Database: EST:\*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1891.720000

score\_list:

| Sequence          | Strd | Orig   | ZScore | EScore  | Len  | Documentation                    |
|-------------------|------|--------|--------|---------|------|----------------------------------|
| gb_est2:BI308442  | +    | 310.00 | 570.07 | 1.7e-22 | 802  | BI308442 EST529852 GP0D Medicag  |
| gb_est2:BG366239  | +    | 284.00 | 520.18 | 1.0e-19 | 809  | BG366239 HVSMEL0006F21f Hordeum  |
| gb_est1:AW720554  | +    | 275.00 | 500.12 | 1.0e-18 | 869  | AW720554 LJNEST1059rc Lotus jaf  |
| gb_est1:AL538252  | +    | 264.00 | 480.01 | 1.7e-17 | 952  | AL538252 LTI_FL013.FB1           |
| gb_est2:BF622268  | +    | 263.00 | 479.35 | 1.9e-17 | 854  | BF622268 HVSMEL0002G16f Hordeum  |
| gb_est1:AL555468  | +    | 259.00 | 470.74 | 5.7e-17 | 927  | AL555468 AL555468 LTI_NFL006.PL  |
| gb_est2:BI179230  | +    | 255.50 | 467.16 | 9.0e-17 | 707  | BI179230 EST520175 cSTE Solanum  |
| gb_est2:BI262982  | -    | 254.00 | 464.42 | 1.3e-16 | 699  | BI262982 NF006E11R1F1086 Irrad   |
| gb_est1:AL515757  | +    | 253.50 | 458.32 | 2.8e-16 | 1092 | AL515757 AL515757 LTI_NFL011.N   |
| gb_est2:BM457834  | +    | 253.00 | 459.98 | 2.2e-16 | 870  | BM457834 AGENCOURT_6412078 NIH   |
| gb_est2:BG325921  | +    | 249.00 | 455.25 | 4.1e-16 | 675  | BG325921 NF084F12ST1F1096 Devel  |
| gb_est2:BF618750  | +    | 249.00 | 453.46 | 5.2e-16 | 788  | BF618750 HVSMEL0007K11f Hordeum  |
| gb_est1:AV925031  | +    | 243.50 | 446.65 | 1.2e-15 | 571  | AV925031 AV925031 K. Sato unpub  |
| gb_est2:BG405336  | +    | 243.50 | 446.00 | 1.4e-15 | 604  | BE405336 WHE1212.E01_J022S Whea  |
| gb_est1:AV916129  | +    | 243.50 | 445.90 | 1.4e-15 | 609  | AV916129 AV916129 K. Sato unpub  |
| gb_est1:AV924234  | +    | 243.50 | 445.66 | 1.4e-15 | 622  | AV924234 AV924234 K. Sato unpub  |
| gb_est1:AV923285  | +    | 243.50 | 445.62 | 1.4e-15 | 624  | AV923285 AV923285 K. Sato unpub  |
| gb_est1:AV924053  | +    | 243.50 | 445.06 | 1.5e-15 | 655  | AV924053 AV924053 K. Sato unpub  |
| gb_est2:BG127495  | +    | 243.50 | 443.42 | 1.9e-15 | 755  | BG127495 EST473141 tomato shoot  |
| gb_est2:BG310222  | +    | 243.50 | 442.17 | 2.2e-15 | 842  | BG310222 HVSMEL0016116f Hordeum  |
| gb_est2:BG470565  | +    | 242.50 | 445.16 | 1.5e-15 | 550  | BE470565 WHE0261.D02_D022S Whea  |
| gb_est2:BG490756  | +    | 242.50 | 444.87 | 1.6e-15 | 564  | BE490756 WHE0368.G11_M222S Whea  |
| gb_est2:BF630423  | +    | 241.50 | 440.19 | 2.8e-15 | 717  | BF630423 HVSMEL0009F05f Hordeum  |
| gb_est2:BF6258379 | +    | 241.50 | 439.67 | 3.0e-15 | 750  | BF258379 HVSMEL0015J02f Hordeum  |
| gb_est2:BF631030  | +    | 241.50 | 439.40 | 3.2e-15 | 768  | BF631030 HVSMEL0014M07f Hordeum  |
| gb_est2:BG26813   | +    | 241.50 | 437.63 | 4.0e-15 | 895  | BF626813 HVSMEL0001G12f Hordeum  |
| gb_est2:BG468478  | +    | 241.00 | 437.14 | 4.2e-15 | 860  | BG468478 G02510452F1 NIH_MGC_15  |
| gb_est1:AV927055  | +    | 240.50 | 441.23 | 2.5e-15 | 555  | AV927055 AV927055 K. Sato unpub  |
| gb_est2:BF628160  | +    | 238.50 | 438.44 | 3.6e-15 | 507  | BF628160 HVSMEL0003E18f Hordeum  |
| gb_est2:BG882033  | +    | 238.50 | 437.82 | 3.9e-15 | 535  | BG882033 sae95b10.y1 Gm-cl065 G  |
| gb_est2:BG452857  | +    | 238.50 | 435.44 | 5.2e-15 | 658  | BG452857 NF081E08LFF1056 Devel   |
| gb_est2:BG298930  | +    | 237.50 | 431.53 | 8.6e-15 | 782  | BG298930 HVSMEL0021E24f Hordeum  |
| gb_est1:AL536791  | +    | 237.00 | 429.09 | 1.2e-14 | 889  | AL536791 AL536791 LTI_FL013.FB1  |
| gb_est2:BF474366  | +    | 236.00 | 433.79 | 6.5e-15 | 501  | BF474366 WHE0843.B11_D212S Whea  |
| gb_est2:BF404585  | +    | 236.00 | 433.57 | 6.7e-15 | 511  | BE404585 WHE0445.A06_A112S Whea  |
| gb_est2:BE9591022 | +    | 236.00 | 433.57 | 6.7e-15 | 511  | BE9591022 WHE1655-1658 J04_J042S |
| gb_est2:BF340158  | +    | 236.00 | 433.34 | 6.9e-15 | 521  | BE340158 EST344227 potato stolic |
| gb_est2:BG045081  | +    | 236.00 | 430.65 | 9.7e-15 | 658  | BG045081 saa38c02.y1 Gm-cl059 G  |
| gb_est2:BF1719579 | +    | 235.50 | 430.01 | 1.1e-14 | 640  | BF1719579 103104503.y1 C. reinh  |
| gb_est2:BF527119  | +    | 235.50 | 428.82 | 1.2e-14 | 601  | BF527119 102407850.y1 C. reinh   |

gb\_est1:AL5559639 + 234.00 423.44 2.4e-14 882 | AL5559639 AL5559639 LTI\_FL011\_

gb\_est1:AW773806 + 233.50 425.86 1.8e-14 658 | AW773806 EST332792 KV3 Medic

gb\_est2:BI417222 + 232.50 426.04 1.7e-14 549 | BI417222 LJNEST1605r Lotus j

gb\_est2:BI418097 + 232.50 424.54 2.1e-14 625 | BI418097 LJNEST24e8r Lotus j

gb\_est1:AU083977 + 232.00 426.48 1.7e-14 486 | AU083977 AU083977 Cryptomeri

seq\_name: gb\_est2:BI308442

seq\_documentation\_block:  
LOCUS BI308442 802 bp mRNA linear EST 20-JUL-2001  
DEFINITION EST529852 GP0D Medicago truncatula cDNA clone pgPOD-5L8 5' end,  
mRNA sequence.

ACCESSION BI308442  
VERSION BI308442.1 GI:14982769  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.

REFERENCE 1 (bases 1 to 802)  
AUTHORS Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho  
J. and Fraser,C.M.

TITLE ESTs from developing reproductive tissues of Medicago truncatula  
JOURNAL Unpublished (2001)

COMMENT USDA/ARS Children's Nutrition Research Center  
Baylor College of Medicine  
1100 Bates Street, Houston, TX 77030-2600, USA  
Tel: 713-798-7044  
Fax: 713-798-7078

Email: mgrusak@bcm.tmc.edu

B395297e

TIGR sequence name: MTOAP64TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA gaa cta gta gat CC).

Location/Qualifiers

source

1..802

/organism="Medicago truncatula"

/cultivar="A17"

/db\_xref="taxon:3880"

/clones="pgPOD-5L8"

/clone\_lib="GP0D"

/tissue\_type="immature pod walls"

/dev\_stage="immature pods, ranging in age from 15 to 30

days after pollination"

/note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Immature pods, ranging in age from 15 to 30 days

after pollination, were collected from greenhouse-grown

plants. At harvest, seeds were removed from pods and

isolated pod walls were collected and immediately frozen

in liquid nitrogen. Pod walls were pooled for mRNA

extraction. cDNA was prepared from polyA+ enriched RNA.

The cDNA was directionally ligated into the Unizap XR

vector from Stratagene and packaged using Gigapack III

Gold packaging extracts. plasmids containing cDNA inserts

were excised from the recombinant lambda-Zap phage using

Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 200 a 157 c 196 g 249 t

ORIGIN

alignment\_scores:

Quality: 310.00 Length: 292

Ratio: 1.813 Gaps: 7

Percent Similarity: 58.562 Percent Identity: 29.110

alignment\_block:

US-09-786-474-2 x BI308442

Align seg 1/1 to: BI308442 from: 1 to: 802

```

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
33 ATGTGGGAATCTTACGATGTTGGGTGGTGCAGACAATCTCAGGCCAA 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 eValProAlaLeuGluAlaLeuProCysMetArgHisArgGlyProA 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 ACCGCTCGCATCATCAATGTCTCGTAGTGGCAGCATAGAGGTCCTG 132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 sAspAlaGlyThrTrpHisAspAlaAlaPheGlyPheAsnArg 50
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 ATGGAGTGGTTGGTCATGTGCATCAAGATGTTACCTCGCTCATCAACGT 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 CTTGCTATTGTGACCTACTCTCCGAGATCAACCTCTT..... 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA 84
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 .TACAACATATGACAACCGTATTATCTACCGTAAATGGGAGATATACA 270
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 ACCACAAGAAATTGAGCAGAACTGAA.....TCTCATCAATTCTGA 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ThrSerGlyAspGlyGluProIleValValGlyPheHisHisTrpGlyGI 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 ACTGGTAGTGACTGTGAAGTCATTGCTCATCTTTATGAAGAATATGGTGA 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 AGAGTTTGTGATATCTCGATGGAGATGTTCTCATTTGTGCTTCTTGATA 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
415 CTGAGATAAAAGTTTCATGTCGTCGTCGATGCTATTGGCATACCCCT 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 LeuPheTyrAlaThrThrGluHisGlyThrVal...PheSerSerGluTy 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
465 CTTTACTTGGTGGGGCGCATGATGATCAATATGTTTTCGCTCGTGAAT 514
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 sLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuA 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
515 GAAGCTCTGATGAGCATGTGAG..... 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 sPlysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluPro 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 ..... 557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 AspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaTh 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 ..... 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 rValArgProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnP 233
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
575 CAGCAACAGGAGGA.....TTAAGAAGGTGGTATAATCCACCAT 615
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 heProValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArg 249
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
616 GGTTCACAGAGAAATTCATCAACAGCTATGATCCACGGTT..... 659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 IleAlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspVa 266
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 TTGCGTGGAGCCCTTTGAGAGGCTGTAGTTAAGAGAATGATGACTGATGT 709
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 lThrValGlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlap 283
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 ACCCTTGGAGTCTTTTCTCCGAGGATGGACTCATCATCTGTTGCTG 759
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 roLeuAlaLysArgHisAsnProAsp 291
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 CAGTGGCCAATCGATATTGGCTGAT 785
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est2.BG366239

```

```

seq_documentation_block:
LOCUS BG366239 809 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEI0006F21f Hordeum vulgare 20 DAP spike EST library HVCNA0010
(20 DAP) Hordeum vulgare cDNA clone HVSMEI0006F21f, mRNA sequence.
ACCESSION BG366239
VERSION BG366239.2 GI:16324128
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 809)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
,R.D., Close,S.J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Morex 20 DAP spike cDNA library
JOURNAL Unpublished (2001)
COMMENT On Mar 8, 2001 this sequence version replaced gi:13255338.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 585
Seq primer: AATTAAACCTCCTAAAGG
High quality sequence start: 6
High quality sequence stop: 730.
FEATURES
Location/Qualifiers
1..809
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEI0006F21f"
/HVCNA0010 (20 DAP)
/tissue_type="20 DAP spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Plants were grown in the greenhouse at the University of
California, Riverside (Fenton, SJ Close, TJ Close). Whole
spikes with awns trimmed were collected at 20 DAP (Fenton
). Total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 1 million
pfu were in vivo excised to give pBluescript SK(-) cDNA
phagemids in the TJ Close lab at the University of
California, Riverside (Choi). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgp/31/cover.html)"
BASE COUNT 173 a 222 c 231 g 182 t 1 others
ORIGIN

```

```

alignment_scores:
Quality: 284.00 Length: 288
Ratio: 1.701 Gaps: 13
Percent Similarity: 57.986 Percent Identity: 31.597

```

```

278 erThraIaIleAala 282
||||: |||
792 CATCACTGGTGCA 805

seq_name: gb_est1:AW720554

seq_documentation_block:
LOCUS AW720554 869 bp mRNA linear EST 19-APR-2000
DEFINITION LJN8T10b9rc Lotus japonicus nodule library 5 and 7 week-old Lotus japonicus cDNA 5', mRNA sequence.
ACCESSION AW720554.1 GI:7615100
VERSION AW720554
KEYWORDS EST.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 869)
AUTHORS Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
TITLE Lotus japonicus root nodule ESTs: tools for functional genomics
JOURNAL Unpublished (2000)
COMMENT Contact: Udvardi MK
Molecular Plant Nutrition
Max Planck Institute of Molecular Plant Physiology
Am Muehlenberg 1, 14476 Golm, Germany
Fax: 49 331 567 8250
Email: udvardi@mpimp-golm.mpg.de
Seq primer: T7
High quality sequence stop: 869.
Location/Qualifiers
1..869
/organism="Lotus japonicus"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/clone_lib="Lotus japonicus nodule library 5 and 7 week-old"
/dev_stage="5 and 7 week-old plants"
/notes="Organ: Nodule; Vector: pSPORT1; Site_1: Salt; Site_2: NotI; The library was prepared using mRNA extracted from nodules of 5 and 7 week-old Lotus plants. Nodules were induced by, and contained Mesorhizobium strain R7A."

BASE COUNT 203 a 182 c 191 g 293 t
ORIGIN

alignment_scores:
Quality: 275.00 Length: 286
Ratio: 1.637 Gaps: 8
Percent Similarity: 58.741 Percent Identity: 28.322

alignment_block:
US-09-786-474-2 x AW720554 ..

Align seg 1/1 to: AW720554 from: 1 to: 869

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla.. 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 ATGTGTGGCATACTTGCTGCTGGTGTCTGATGACTCTCAAGCCAA 189

17 .....PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 AAGGGTTCGCGCTCCGTGAGTCTTCTCGCAGATTG.....AAGCACC 230

31 rGglyProAspAlaGlyThrTTPHisAspAlaAspAlaAlaPheGly 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 GTGGCCCTGACTGGAGTGGCTGCCAACACATGGTGATTAACCTTTTGGCT 280

48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

281 CATCAACGTTTACAAATAGTTGATCCAGCTTCGGTGATCAACACTCCTC... 328
64 gTgLyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
   :||||| :||| :|||||
329 .....TTCAATGAAGATCAATCTATCTATGTCACGGTGAATGGAG 368
81 luTleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
   :||||| :||| :|||||
369 AGATATTCAATCATGAAGAACTCAGGAACAATTCGCAAAAT.....CAC 412
98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisH 114
   :||||| :||| :|||||
413 AAGTTTCAGAACGATGTGATGATGATGATGATGATGATGATGATGATG 462
114 sTgPlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
   :||||| :||| :|||||
463 GCATGGAGAAACTTGTGGACATGCTGGATGATGATGATGATGATGATG 512
131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
   :||||| :||| :|||||
513 TGCTGGACACCCGCGACAACAGTTTTTTAGTTCGAAGGATGCTATAGGG 562
148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
   :||||| :||| :|||||
563 GTCACCTCCCTTATACATTTGTTATGCACTAGATGGGTCCGTTTGGATT 612
163 rSerGluLysLysThrIleLeuGluMetAlaGluMetAsnLeuAspL 180
   :||||| :||| :|||||
613 ATCAGAATTGAAGGGCTGAATGATGATGATGATGATGATGATGATGAT 661
180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
661 ..... 661
197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
   :||||| :||| :|||||
662 ...CCACCCGGTCACCTGTACTCTAGCAAGGAAAAA..... 694
213 sThrAlaThrValArgProGlyLysLeuGluGlnLysArgTyrPheL 230
   :||||| :||| :|||||
695 .....GAGTCCGTAGATGATGATGATGATGATGATGATGATGATGAT 713
230 ysProGlnPheProValGlnLysValValLysGlyLysGluGlnAspLeu 246
   :||||| :||| :|||||
714 ACCCTCATGGTCTCTGAGGCTATTCATCAGCCCTTATGATCCACTA 763
247 PheAspArgIleAlaGlnValLeuGluAspSerValGluLysHisMetAr 263
   :||||| :||| :|||||
764 .....GCTTTGAGGAGGAGGCTTTGAAGAAGGCTATCATAAAGGTTGAT 807
263 gAlaAspValThrValLysSerPheLeuSerGlyGlyIleAspSerThrA 280
   :||||| :||| :|||||
808 GACTGATGTGCTTTTGGAGTTCTATTGTTGAGGTTTGGAGTTTGGACTCT 857
280 laIleAla 282
858 TGTTGCA 865
seq_name: gb_est1:AL538252

```

```

seq_documentation_block:
LOCUS AL538252 952 bp mRNA linear EST 13-FEB-2001
DEFINITION AL538252 LTI_FL013_FBrnl Homo sapiens cDNA clone CSODF027Y113 5
prime, mRNA sequence.
ACCESSION AL538252
VERSION AL538252.1 GI:12801745
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 952)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

```

TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
FEATURES  
source  
1..952  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="LTI\_FL013\_FBrnl"  
/dev\_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"  
/lab\_host="DH10B"  
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(df) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 260 a 199 c 228 g 262 t 3 others  
ORIGIN

alignment\_scores:  
Quality: 264.00 Length: 325  
Ratio: 1.509 Gaps: 15  
Percent Similarity: 53.846 Percent Identity: 27.385

alignment\_block:  
US-09-786-474-2 x AL538252 ..

Align seg 1/1 to: AL538252 from: 1 to: 952

```

19 ProAlaLeuGluArgAlaLeuProCysMetArg.....HisArgGlyPr 33
   :||||| :||||| :|||||
84 CCGTCTACGCCAGTGTCTGAGTGTCTATGAGATGTCACACAGAGTCC 133
   :||||| :||||| :|||||
33 oAspAspAlaGlyThrTrpHisAsp.....AlaAspAlaAlaP 46
   :||||| :||||| :|||||
134 A...GATGCATTCGGTTTGGAGAAATGTCATGATGATGATGATGATGAT 180
   :||||| :||||| :|||||
46 heGlyPheAsnArgLeuSerIleAlaAspIleAlaHisSerHisGlnPro 62
   :||||| :||||| :|||||
181 TTGGATTTCACCGTTTGGCGGTAGTTGACCCGCTGTTTGGAAATGCAGCCA 230
   :||||| :||||| :|||||
63 LeuArg.....TrpGlyProAlaAspGluProAs 72
   :||||| :||||| :|||||
231 ATTCGAGTGAAGAAATATCCGTATTTCGG..... 260
72 pArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrValGluLeuA 89
   :||| :||| :|||
261 .....CTCTGTACAAATGGTGAATCTACACCAATGAAGATGC 300
89 rgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGlyAspGly 105
   :||||| :||||| :|||||
301 ACACGATTTTGA.....TTTGAATACACGACCAAAAGTGGATGGT 341
106 GluProIleValValGlyPheHisHisTrpGly...GluSerValValG 121
   :||| :||| :|||
342 GAGATAATCCTTCATCTTATGACAAAGGAGGAATGACCAACAATTCG 391
121 uHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGluLys 138
   :||| :||| :|||
392 TATGTTGGAGTGTGTTGTCATTTGTTTACTGATGATGATGATGATGAT 441
138 erLeuPheLeuAlaArgAspGlnPheGlyIleLysProLeuPheTyrAla 154
   :||| :||| :|||

```

```

442 AAGTGTTCCTGGGTAGATACATATGAGTCAGACCTTTGTTTAAAGCA 491
155 ThrThrGluHisGly...ThrValPheSerSerGluLysLysThrIleLe 170
492 ATGACAGAAGATGATTTGGCTGTATGTTTCAGAAGCTAAAGCTGTGT 541
170 uGluMetAlaGluMetAlaLeuAspLeuGlyLeuAspLysArgThrI 187
542 TACATTGAAGCAGCTCCGCGACTCCTTTTAAAGTGAGGCTTTCTTC 591
187 leGluHisTyrValAspLeuGlnTyrValPro..... 197
592 CTGACACATATGAAGTTTGGATTAAAGCCCAATGGCAAGTTGCATCC 641
198 .....GluProAspThrLeuHisAlaG1 205
642 GTGGAATGGTTAATATCATCATCTGCGGATGAACCCCTGCACGCCCT 691
205 nileSerArgLeuGluSerGlyCysThrAlaThrValArgProGlyGlyL 222
692 CTATGACAATGTGGAGAA.....CTCTTTCCAGGTTTGT 726
222 ysLeuGluGlnLysArgTyrPhelysProGlnPheProValGlnLysVal 238
727 AGATAGAA.....ACT 737
239 ValLysGlyLysGluGlnAspLeuPheAspArgIleAlaGlnValLeuG1 255
738 GTGAAGAACACCTCAGGATCCTTTTAAAT ..... 767
255 uAspSerValGluLysHisMetArgAlaAspValThrValGlySerPheL 272
768 .AATGCTCTAAGAACACCTTTGATGACAGACAGAGGATTGCTCCCTTT 816
272 euSerGlyGlyIleAspSerThrAlaIleAla.ProLeuAla..... 285
817 TATCAGGGGGTGTGACTCCAGCTTGGTTGCTGCCACTCTGTGAAGCAG 866
286 .....LysArgHis..... 288
867 CTGAAGAAGCCCAAGTACATATCTCTCCAGACATTGCAATTGGCAT 916
289 .....AsnProAspLeuLeu 293
917 GGAAGAMAGCCCGGATTTACTG 939
q_name: gb_est2:BF622268

```

```

seq_documentation_block:
LOCUS BF622268 854 bp mRNA linear EST 17-OCT-2001
DEFINITION HVSMEd0002G16f Hordeum vulgare seedling shoot EST library
HVCdNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEd0002G16f,
mRNA sequence.
ACCESSION BF622268
VERSION BF622268.2 GI:13080047
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 854)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Yu
,X., Henry,B., Palmer,M., Rambo,T., Simmons,J., Oates,R., Choi,D.W.,
Fenton,R.D. and Main,D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Morex cold-stressed seedling shoot cDNA
library
JOURNAL Unpublished (2001)
COMMENT On Dec 18, 2000 this sequence version replaced gi:11886002.
Contact: Wing RA
Clemson University
Clemson University

```

```

100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total hg bases = 574
Seq primer: AATTAAACCTCACTAAAGGG
High quality sequence stop: 841.
Location/Qualifiers
1. .854
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEd0002G16f"
/vclone_lib="Hordeum vulgare seedling shoot EST library
HVCdNA0001 (Cold stress)"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 50C for 2 days. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 600000 pfu
were in vivo excised to give pBluescript SK(-) cDNA
phagemids. These steps were performed in the TJ Close
laboratory at the University of California, Riverside
(Choi, Close, Fenton). Phagemids were plated and picked at
the Clemson University Genomics Institute (CUGI) (Begum,
Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations
, DNA sequencing and sequence analysis were performed at
CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main
). The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)*
BASE COUNT 191 a 232 c 235 g 196 t
ORIGIN

```

## alignment\_scores:

```

Quality: 263.00 Length: 279
Ratio: 1.594 Gaps: 13
Percent Similarity: 59.140 Percent Identity: 30.824
alignment_block:
US-09-786-474-2 x BF622268 ..
Align seg 1/1 to: BF622268 from: 1 to: 854
1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGlu..... 15
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 ATGTGTGGCATACTGTCAGTGTGGCTGCGCCGATGACACCCAGGGGAA 106
16 ....AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 GAGAGTGGGGTGTCTGACGCTCTCGCGCAGGCTC.....AAGCACC 147
31 rgGlyProAspAlaGlyThrTrpHisAspAlaAspAlaPheGly 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
148 GCGGCCCCGACTGGAGTGGCATGCACCGAGTGGGTGACTACCTCTCC 197
48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
198 CACCAGCGCTCGCCATCATCGACCTGCTCAGGCGACCGCCACTT.. 245

```

```

64  gTgPlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
246  .....TACAACGAGCACAAGTCCATCGTCGTCACGTGAATGGAG 285
81  lulleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
286  AGATCTACAACCATGAACAGCTCGGCGCAGAGCTCC.....TCCAC 329
98  ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisH 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330  ACATTCAGGACAGCAGCAGCTCGAGGTCTATCGCACACCTGTACGAGGA 379
114  sTgPlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
380  GCATGGGAGCAATTCATCGCATGCTGGATGCTGCTCTCTCCTTCGTCT 429
131  letrPaspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430  TGCTCGACACACGCGACACAGCTTCATTGCTGCCGCTGATGCCATTGGT 479
148  ileLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480  GTCACACCCATGATGTGTGCTGGGNAATGATGGTGGTGGTGGATATC 529
163  rSerGluLysLysThrIleLeuGluMetAlaGluMetAsnLeuAspL 180
||||| |||
530  ATCAGAGATGAAG..... 542
180  euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543  .GGCCTGAATGAT...GATTGTGAGCAGCTTT.....GAGATCTTT 578
197  ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
||| ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
579  CCA...CCTGCTCATCTCTACTCCAGC..... 602
213  sThrAlaThrValArgProGlyGlyLysLeuGluGlnLysArgTyrPheL 230
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
603  .....AAGGAGGGAGGC.....TTCAAGAGATGGTACA 630
230  ysProGlnPheProValGlnLysValValLysGlyLysGluGlnAspLeu 246
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
631  ACCACCTGG...GTCGCGAGGTCAATTCCTCAGTGCCATATGATCCA 677
247  PheAspArgIleAlaGlnValLeuGluAspSerValGluLysHisMetAr 263
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
678  CTT...GCTCTCAGGAAGCTTTTCGAATAGGCTGTACCAAGAGGCTTAT 724
263  gAlaAspValThrValGlySerPheLeuSerGlyGly 275
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725  GACAGAGCTTCCATTCGGTGTCTTACTCTTTGGTGGG 761

```

\_name: gb\_est1:AL555468

```

seq_documentation_block:
LOCUS      AL555468                      927 bp    mRNA    linear    EST 16-FEB-2001
DEFINITION AL555468 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK008YP10 5
            prime, mRNA sequence.
ACCESSION  AL555468
VERSION    AL555468.1 GI:12897229
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 927)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

```

# FEATURES

```

source
Location/Qualifiers
1..927
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 257 a 178 c 213 g 277 t 2 others
ORIGIN
alignment_scores:
Quality: 259.00 Length: 316
Ratio: 1.515 Gaps: 13
Percent Similarity: 54.114 Percent Identity: 26.266
alignment_block:
US-09-786-474-2 x AL555468 ..
Align seg 1/1 to: AL555468 from: 1 to: 927
2 CysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPheVa 18
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
90 TGACCATGTGTGGCATTTGGCGCCTTTTGGCAGTCATGATTCCTT... 137
18 lProAlaLeuGluArgAlaLeuProCysMet.....ArgH 30
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 .....TCTGTCAGTCTCTGAGTGTGCTATGAAGATTGCAC 171
30 isArgGlyProAspAspAlaGlyThrTrpHisAsp.....Ala 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 ACAGAGTCCA...GATGCATTCGGTTTGGAGATGTCAATGGATACACC 218
43 AspAlaAlaPheGlyPheAsnArgLeuSerIleIleAspIleAlaHisSe 59
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 AACTGCTGCTTTGGATTTTCATGCTGGCGGTAGTTGACCCGCTGTTGG 268
59 rHisGlnProLeuArg.....TrpGlyProAlaA 69
:|||||:|||||:|||||:|||||:|||||:|||||:
269 AATGCAGCCCAATTCGAGTGAAGAAATATCCGTATTGTGTG..... 308
69 spGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyr 85
:|||||:|||||:|||||:|||||:|||||:|||||:
309 .....CTCTGTTACATGTTGAATCTACACCAT 338
86 ValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSe 102
:|||||:|||||:|||||:|||||:|||||:|||||:
339 AACAGAGTCAACAGCAGCATTTTGAA.....TTTGAATACCAAGACCAA 379
102 rGlyAspGlyGluProIleValValGlyPheHisHisTrpGly...GluS 118
:|||||:|||||:|||||:|||||:|||||:|||||:
380 AGTGGATGGTGAGATAATCCTTCATCTTTATGACAAAGAGGAATTGAGC 429
118 erValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThr 134
:|||||:|||||:|||||:|||||:|||||:|||||:
430 AACATTTGTATGTGGATGGTGTGTTGCATTTGTTTACTGGATACT 479
135 LysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysProLe 151
:|||||:|||||:|||||:|||||:|||||:|||||:
480 GCCAATAAGAAAGTGTCTCGGTAGAGATACATATGAGTTCAGACCTTT 529
151 uPheTyrAlaThrThrGluHisGly...ThrValPheSerSerGluLysL 167
||||| ||| ||||||| ||| :|||:|||||:|||||:

```



```

530 GTTTAAACAATGACAGAGATGGATTTTGGCTGTATGTTTCAGAGCTA 579
167 ysThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAsp 183
|| : : : : : : : : : : : : : : : : : : : : : : : : :
580 AAGGTCCTGTTACATGAAGACACTCGGACTCCCTTTTAAAGTGGAG 629
184 LysArgThrIleGluHisTyrValAspLeuGlnTyrValPro..... 197
||||| : : : : : : : : : : : : : : : : : : : : :
630 CCTTTCTCTCGGACACTATGAAGCTTTGGATTTAAAGCCAAATGGCAA 679
198 ..... : : : : : : : : : : : : : : : : : : : :
680 AGTGCATCCTGGAAATGGTTAAATATCATCATCTCGGGATGAACCC 729
202 euHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValArg 218
||||| : : : : : : : : : : : : : : : : : : : :
730 TGCAGCCCTCTATCAATGTGGAGAA.....CTCTTT 764
219 ProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProVa 235
||||| : : : : : : : : : : : : : : : : : : : :
765 CCAGCTTTTGAGATAGAA..... : : : : : : : : : : : : : :
235 iClnLysValLysGlyLysGluGlnAspLeuPheAspArgIleAlaG 252
783 .....ACTGTGAARAACACCTCAGGATCCTTTTAA..... 815
252 lnValLeuGluAspSerValGluLysHisMetArgAlaAspValThrVal 268
: : : : : : : : : : : : : : : : : : : : : : : : :
816 .....AATGCTCTAAGAAACGTTTGATGACAGACAGAAGATT 854
269 GlySerPheLeuSerGlyGlyLeuAspSerThrAlaIleAlaProLeu 284
||| : : : : : : : : : : : : : : : : : : : :
855 GGCTCCCTTTATCAGGGGCTTGACTCCAGCTGGTGGTGTGCACCTC 902

```

seq\_name: gb\_est2:BI179230

```

seq_documentation_block:
LOCUS      BI179230               707 bp    mRNA    linear    EST 09-JUL-2001
DEFINITION EST520175 cSTE Solanum tuberosum cDNA clone cSTE17F18 5' sequence,
            mRNA sequence.
ACCESSION  BI179230
VERSION    BI179230.1 GI:14645041
SOURCE     EST.
ORGANISM   Solanum tuberosum
            potato.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 707)
AUTHORS   van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J.,
            Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and
            Baker,B.
            Generation of ESTs from in vitro grown microtubers
            Unpublished (2001)
            Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13P-R.

```

```

FEATURES             Location/Qualifiers
     source            1..707
                     /organism="Solanum tuberosum"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTE17F18"
                     /clone_lib="cSTE"
                     /tissue_type="axillary buds of stem explants; growing
                     sink-tubers"
                     /dev_stage="7, 8 and 10 days"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Tissue supplied by Christian Bachem and Richard
                     Visser (Department of plant Breeding, Wageningen

```

University, The Netherlands). The cSTA libraries will attempt to capture the induction and initiation/initial growth of the tuber in an in vitro system as described in Bachem et al. (Plant Journal, 1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, cSTA (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, cSTA (21-40) and cSTA (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as P3 in Tanksley lab notebooks."

BASE COUNT 191 a 133 c 168 g 215 t  
ORIGIN

alignment\_scores:  
Quality: 255.50 Length: 175  
Ratio: 2.147 Gaps: 3  
Percent Similarity: 68.000 Percent Identity: 35.429

alignment\_block:

US-09-786-474-2 x BI179230 ..

Align seg 1/1 to: BI179230 from: 1 to: 707

```

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaph 17
||||| : : : : : : : : : : : : : : : : : : : :
11 AVGTGTGGAATACTTGCAATTTTCGGTTGCCTGATTAATTCATGCCAA 60
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
: : : : : : : : : : : : : : : : : : : :
61 GCGTTCAAGAATCATCGAACTATCAAGAAGTTGCCCATAGAGGACCTG 110
34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg 50
|| : : : : : : : : : : : : : : : : : : : :
111 ATTGGAGTGGATGTGATAGCCATGAGGACTGTATCTGCTCATCAACGA 160
51 LeuSerIleAlaAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67
||||| : : : : : : : : : : : : : : : : : : : :
161 TTGGCAATAGTAGACCCCAACTTCAGGGGATCAGCCGCTG..... 199
67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA 84
: : : : : : : : : : : : : : : : : : : :
200 .TATATGAGGACAGACACCATTTGTTGTCGGTAAATGGAGAGATCTACA 248
84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
|| : : : : : : : : : : : : : : : : : : : :
249 ACCATAAGGAATACGGGAGAAACTGAAG.....TCTCATCAGTTTCGA 292
101 ThrSerGlyAspGlyGluProIleValGlyPheHisHisTrpGlyG 117
||||| : : : : : : : : : : : : : : : : : : : :
293 ACTGAAGTGGATTTGGAAGTTATGCCCATCTTTATGAAGAATATGGAGA 342
117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAsp 134
|| : : : : : : : : : : : : : : : : : : : :
343 AAATCTCATTTGACATGTGGATGGGATGTTCTTTTGTCTTCTTGATA 392
134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150
||||| : : : : : : : : : : : : : : : : : : : :
393 CCCGGGATAAAAGTTTCATTGCTGCTCGGGATGCCATTGGCATACACCC 442
151 LeuPheTyrAlaThrThrGluHisGlyThrVal...PheSerSerGluLys 166
||||| : : : : : : : : : : : : : : : : : : : :
443 CTTTATATGGGTGGGGTCTTCATGGTCCCATATGGTTTCTCTCAGAGAT 492
166 slsYThrIleLeuGluMetAlaGlu 174
|| : : : : : : : : : : : : : : : : : : : :
493 GAAGCCCTTAAGTGATGATGTGAA 517

```

seq\_name: gb\_est2:BI269282





JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LLAM12131 row: 1 column: 20

High quality sequence stop: 687.

## FEATURES

```

/organism="Homo sapiens"
/ab_xref="taxon:9606"
/clone_image="5499043"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC library."

```

BASE COUNT  
ORIGIN

162 c 211 q 260 t

alignment\_scores:

|                     |        |                   |        |
|---------------------|--------|-------------------|--------|
| Quality:            | 253.00 | Length:           | 314    |
| Ratio:              | 1.497  | Gaps:             | 13     |
| Percent Similarity: | 53.822 | Percent Identity: | 26.115 |

alignment block:

US-09-786-474-2 x BM457834

Align seq 1/1 to: BM457834 from: 1 to: 870

```

2  CysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPheVa 18
   |||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
22  TGCACCATGCTGGGCATTTGGCGCGTGTGGCCAGTGATGATTGCCCTT.. 69
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
18  lProAlaLeuGluArgAlaLeuProCysMet.....ArgH 30
   ::  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
70  ....TCTGTCAAGTCTCAGTGCTCATGAAGATTGCAC 103
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
30  isArgGlyProAspAlaGlyThrTrpHisAsp.....Ala 42
   |||||  |||||  |||||  |||||  |||||  |||||  ::  ::
104  ACAGAGGTCCA...GATGCATTCCTTTTGAGATGTCATGGATACACC 150
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
43  AspAlaAlaPheGlyPheAsnArgLeuSerIleIleAspIleAlaHis 59
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
151  AACTGCTGCTTTGGATTTTCACCGGTTGGCGGTAGTTGACCGGTGTTGG 200
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
59  rHisGlnProLeuArg.....TrpGlyProAlaA 69
   ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
201  ATGCAGCCCAATTCAGGTGAGAAATATCCGATTGTGG..... 240
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
69  spGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyr 85
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
241  ....CTCTGTTACATGCTGGAATCTACACCAT 270
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
86  ValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThr 102
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
271  AAGAAGATGCACACCATTTTGA.....TTTGAATACCCAGACAA 311
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
102  rGlyAspGlyGluProIleValValGlyPheHisHisTrpGly..... 118
   ::  |||||  |||||  |||||  ::  ::  ::  ::  ::  ::  ::
312  AGTGGATGGTGAGATAATCCCTTCATCTTTATGACAAGAGGAATTGAGC 361
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
118  ervAlValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThr 134
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

362 AACAAATTTGATGTTGGATGGTGGTGGTGGATTTGCAATTTGTTTACTGGATACT 411

135 LysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysProIle 151  
:::|||||:::|||||:::|||||:::|||||:::|||||

412 GCCAATAAGAAAGTGTCTCGGTAGAGATACATATGCAGTCAGACCTTT 461

151 uPheTyrAlaThrThrGluHisGly...ThrValPheSerSerGluLysL 167  
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||

462 GTTTAAAGCAATGACAGAAGATGGATTTTGGCTGTATGTTCAGAAAGCTA 511

167 ysthIleLeuGluMetAlaGluMetAsnLeuAspLeuGlyLeuAsp 183  
||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||

512 AAGGCTCTGTTCATCTGAAGCCTCCGCCACTCCCTTTTAAAGTGGAG 561

184 LysArgThrIleGluHisTyrValAspLeuGlnTyrValPro..... 197  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

562 CCTTTTCTCTCGACACTATGAAGTTTGTGATTTAAAGCCAAATGGCAA 611

198 .....GluProAspThrL 202  
::: ::::: ||| ::::: ||| ::::: ||| ::::: |||

612 AGTTGCATCCGTGGAAATGGTTAAATATCATCACTGTCGGGATGAACCCC 661

202 euHisAlaGlnIleSerArgLeuGluGlnSerGlyCysThrAlaThrValArg 218  
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||

662 TGCAGCCCTCTATGACAACTGGAGAAA.....CTCTTT 696

219 ProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProVa 235  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

697 CCAGGTTTTCAGATAGAA..... 714

235 lGlnLysValValLysGlyGlyGluGlnAspLeuPheAspArgIleAlaG 252  
:::|||||:::|||||:::|||||:::|||||:::|||||

715 .....ACTGTGAAGAACCACTTCAGATCCTTTT. 744

252 lNValLeuGluAspSerValGluLysHisMetArgAlaAspValThrVal 268  
:::|||||:::|||||:::|||||:::|||||:::|||||

745 .....TATATGCTGTTAAGAAACGTTTGTATGCAGACAGAGAGGATT 786

269 GlycerPheLeuSerGlyIleAspSerThrAlaIleAla 282  
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

787 GCGTCCTTTTATCAGGGGCTTCGACTCCAGCTTGGTTGCT 828

seq name: ab est2: BE325921

```

seq_documentation_block:
LOCUS       BE325921                675 bp      mRNA
DEFINITION  NF084F12ST1F1096 Developing stem Medicago truncatula cDNA clone
            NF084F12ST 5', mRNA sequence.
ACCESSION   BE325921
VERSION     BE325921
KEYWORDS    EST.
SOURCE      BE325921.1 GI:9199698
            barrel medic.
ORGANISM    Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1 (bases 1 to 675)
AUTHORS    He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            ,R.A.
TITLE       Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
JOURNAL     Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Medicago Genome Initiative accession: MGI:S:26492
            Insert Length: 675 Std Error: 0.00

```

```

202 LeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVal 218
      |||
521 .....CCAGGCATATATATCCAGCAA 543

218 gProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPhePro 235
      |||||||
544 ACAGGGAGGA.....TTAAGAAGGTGGTATATCCACCATGGTTCA 584

235 alGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIleAla 251
      |||||||
585 CAGAGAAATTCATCAACAGCCTATGATCCACGGTT.....TTGCGT 628

252 GlnValLeuGluAspSerValGluLysHisMetArgAlaAspVal 266
      |||||||
629 GAAGCCCTTCGAGGGCGCTAGTTTAAAGAAATGATGACTGATGTA 673

```

seq\_name: qb\_est2:BF618750

|            |   |        |                             |
|------------|---|--------|-----------------------------|
|            | <b>seq_documentation_block:</b>   |        |                             |
| LOCUS      | BF618750  | 788 bp | mRNA linear EST 22-OCT-2001 |
| DEFINITION | HVSMEc0007K1lf Hordeum vulgare seedling shoot EST library<br>HYCDNA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone<br>HVSMEc0007K1lf, mRNA sequence. |        |                             |

|           |  |
|-----------|--|
| ACCESSION | BF618750.2   |
| VERSION   | GI:13107381  |
| KEYWORDS  | EST.   |
| SOURCE    | barley.  |
| ORGANISM  | Hordeum vulgare  |
| REFERENCE | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.                   |
| AUTHORS   | Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Oates, R. and Main, D. |
| TITLE     | Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling shoot cDNA library  |
| JOURNAL   | Unpublished (2001)   |
| COMMENT   | On Dec 18, 2000 this sequence version replaced q1:11882484.  |

Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA

100 JOHNS HALL, CLEMSON  
Tel: 864 656 7288

TEL: 864 656 7288  
FAX: 864 656 4293

FAX: 864 636 4293  
 Email: [rwjng@clmson.edu](mailto:rwjng@clmson.edu)

Email: [rwing@clmson.edu](mailto:rwing@clmson.edu)  
Total pages = 132

Total hq bases - 432

Seq primer: AATTAAACCCCTCACTAAAGGG  
high quality: sequence stop: 750

High quality sequence stop: 750.

| FEATURES            |        |
|---------------------|--------|
| Location/Qualifiers | 1, 700 |

| source | 1. .788 |
|--------|---------|
|--------|---------|

/organism="Hordeum v

/cultivar="Morex"

/db\_xref="taxon:451

/clone="HVSMEc0007K

```
/clone_lib="Hordeum
```

HVC DNA0003 (Etiolat)

/tissue\_type="Seedl:

```
lab_host="TJC121"
```

```

/note="Vector: lambda

```

Seeds were surface sterilized with 10% bleach for 1 min, washed 3 times with sterile distilled water, and then surface-sterilized with 70% ethanol for 30 s.

conditions in the data

conditions in the presence of water, nystatin

crystallization dis(

crystallization  
then harvested, total

purified, one primary

purified, one primary and 1 million pfu were

and 1 million pfu were

CDNA phagemid library SK(-) close laboratory at

Close laboratory at  
piverside / Choi cl

Riverside (Chol, Cl)

picked at the Clemson

(Begum, Palmer, Frie

preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"









**THIS PAGE BLANK (USPTO)**

OM of: US-09-786-474-2 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Sep 20, 2002 9:32 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09786474/runat\_18092002\_132912\_5495/app\_query.fasta\_1.706  
-DB=Issued\_Patents\_NA\_QPWT=fastap -SUFFIX=p2n.rni  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOCPCL=0.000  
-LOOPEXT=0.000 -XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000  
-XGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-NATRX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
OUTWNT=pfs -NORM=ext -HRAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
SER=US09786474@cgn1\_1\_54 -NCPU=6 -ICPU=3 -LONGLOG  
EV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-786-474-2  
Query length: 640  
Database: Issued\_Patents\_NA\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 54.020000

score\_list:  
Sequence Strd Orig zScore EScore Len Documentation  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-125-468-1 633.00 1202.64 2.7e-59 30001  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-474-933-1 633.00 1202.64 2.7e-59 30001  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-09-090-793-12 463.50 856.16 5.3e-40 40138  
/cgn2\_6/ptodata/1/ina/backfile1.seq:5256558-3 425.00 816.83 8.3e-38 2002  
/cgn2\_6/ptodata/1/ina/backfile1.seq:5256558-1 421.00 809.51 2.1e-37 1886  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-737-825-1 184.00 311.18 1.2e-09 8341  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-385-028-16 165.00 294.45 1.0e-08 1542  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-385-028-13 165.00 268.52 2.9e-07 11604  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-385-028-11 165.00 265.15 4.4e-07 15079  
/cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-599-171A-22 154.50 271.38 2.0e-07 1779  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-646-590B-22 154.50 271.38 2.0e-07 1779  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-09-069-226-22 154.50 271.38 2.0e-07 1779  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-412-184-22 154.50 271.38 2.0e-07 1779  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-328-111-544 152.00 280.95 5.8e-08 570  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-328-111-395 132.00 239.25 1.2e-05 629  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-428-517-1 131.50 181.77 0.0195 50937  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-906-925-5 122.00 202.16 0.0014 2337  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-902-633-3 118.50 195.15 0.0035 2325  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-328-111-401 116.50 207.23 0.0007 663  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-804-196-1 116.50 177.11 0.0354 6909  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-658-340-1 116.50 177.11 0.0354 6909  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-746-111-26 116.50 177.11 0.0354 6909  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-949-637-1 115.00 193.78 0.0042 1491  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-09-291-488-1 115.00 193.78 0.0042 1491  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-328-111-668 114.50 203.26 0.0012 659  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-328-111-456 112.00 200.32 0.0018 559  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-911-445-4 111.50 188.72 0.0080 1275  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-182-983-4 111.50 188.72 0.0080 1275  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-911-445-19 111.50 186.99 0.0100 1458  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-182-983-19 111.50 186.99 0.0100 1458  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-911-445-17 111.50 184.31 0.0141 1796  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-182-983-17 111.50 184.31 0.0141 1796  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-911-445-5 111.50 183.97 0.0347 1845  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-08-911-445-6 111.50 183.97 0.0347 1845  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-911-445-6 111.50 182.64 0.0174 2046  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-08-182-983-6 111.50 182.64 0.0174 2046  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-375-134-13 111.00 177.90 0.0320 2734  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-09-057-570-1 111.00 166.32 0.1415 6736  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-09-057-570-5 111.00 166.32 0.1415 6736

/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-09-057-570-3 + 111.00 165.82 0.1509 7004  
/cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-08-952-127-11 + 110.00 159.72 0.3299 9620  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-844-056-1 + 104.50 169.14 0.0985 1944  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-743-637B-6 + 104.50 164.91 0.1695 2703  
/cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-526-840B-6 + 104.50 164.91 0.1695 2703

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-125-468-1

seq documentation block:  
; Sequence 1, Application US/08125468  
; Patent No. 5589385  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: Clontetracycline and tetracycline formation and cosmid  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125,468  
; FILING DATE: 22-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tsevdos, Estelle J  
; REGISTRATION NUMBER: 31,145  
; REFERENCE/DOCKET NUMBER: 31,255-02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201)831-3241  
; TELEFAX: (201)831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30001 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-125-468-1  
  
alignment\_scores:  
Quality: 633.00 Length: 622  
Ratio: 1.829 Gaps: 20  
Percent Similarity: 55.627 Percent Identity: 29.582  
  
alignment\_block:  
US-09-786-474-2 x US-08-125-468-1/rev ..  
  
Align seg 1/1 to reverse of: US-08-125-468-1 from: 1 to: 30001  
  
1 MetCysGlyLeuLeuGlyLeuThrAlaAsnGlyAsnAlaGluAlaPh 17  
|||||.....|:|||||  
22914 ATGTGCGGAATCGTGGGGTGTGCTGACTACGACGCCGCCGACACCG 22865  
  
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34  
|||||.....|:|||||  
22864 CCGGACCTGCTCCAGGCGATGACCGACCATGCGCTCGCCGGGGCCGG 22815  
  
34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg 50  
||| ||| ||| |||:|||||

22814 ACGCCGAGGCGCTCTGTTCCGCGCGCGCGGCTCGGTCCACCGCGC 22765  
51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67  
22764 CTGTCTGATGACCCCGAGCACGCGCGCGACCGCATG.....AC 22724  
67 oAlaAspGluProAspArg...TyrAlaMetThrPheAsnGlyGluIle 83  
22723 CGCGGAACACGAGGCGCGCGCGCGCGATCACCCTTCAGCGGAGATCT 22674  
83 TyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPhe 99  
22673 ACAACTTCGCTGAGCTGCGCGCGCACTCACCCTCGCACGCGCCGTTTC 22624  
100 AsnThrSerGlyAspGlyGluProIleValValGlyPheHisIleTrpG 116  
22623 AGGACCTCTCGATACCGAGGTGCTCTCGCGGCTACTCTCAGTGGG 22574  
116 yGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpA 133  
22573 CGCGCGCTGTGCGAGAGCTCAACGGATGTTGCGCTTCGCGCTCTGG 22524  
133 spThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLys 149  
22523 ACGACACAGCGAGGAACCTCTGCTGCTCGCGACCGGATGGCGTCAAG 22474  
150 ProLeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGly 166  
22473 CGGCTCTACTTGGCGGACCGCGCGGCTGCGCTTCGGCTCGGAGCC 22424  
166 sLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuA 183  
22423 CAAGCGCGTCTG.....GCGGACCGGACCGCTGT 22395  
183 sLysArgThr.....IleGluHisTyrValAspLeuGlnTyr 195  
22394 CCGCGCGGTGCGCGGCGGCTGCGAGGTGCTGGACATGGTGAAG 22345  
196 ValProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerG 212  
22344 ACGCCGAGGCGCGCTCTCTCCGGGCTCGTCGAG..... 22309  
212 yCysThrAlaThrValArgProGly.....GlyL 222  
22308 .....GTCCGTCGGGTACGTCGTCGGGTGCGGCGCGG 22272  
222 ysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnLysVal 238  
22271 GGTGCGCGCGGACGCTACTGCGGCTCGAG..... 22240  
239 ValLysGlyLysGluGlnAspLeu.....PheAspArgIleAlaG 252  
22239 GCCCGGAGCACACCGACGACCTGCGCGACCATCGGACCGTCCGTGA 22190  
252 nValLeuGluAspSerValGluLysHisMetArgAlaAspValThrValG 269  
22189 CCGTCTGCGCGACACCGCTACCGCTGAGTGTCTCGGACGTCCTCGG 22140  
269 LysSerPheLeuSerGlyIleAspSerThrAlaIleAlaProLeuAla 285  
22139 GCAGCTGTCTCCGCGGCTGAGCTGTCGCGGCTGAGCGGCTGGCC 22090  
286 LysArgHisAsnProAspLeuLeuThrPheThrThrGlyPhe..... 299  
22089 GCCCGCAGCGGCGG...GTGCGCTCTCTCTCGGTGACTTCGCGGCGC 22043  
300 .....GluArgGluGlyTyrSerGluV 307  
22042 CGCCGACGCTTCGCGCTCGGACGCGCTCGCGGCGGAG.....GCGAGC 21999  
307 alAspValAlaAlaGluSerAlaAlaIleGlyAlaGluHisIle... 322  
21998 CGCGGTACGTGCGGAGTGGCGGCGACGTCGAGCGGACCACTTGAG 21949

323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe 339  
21948 GTGCTCTGGACAGCCCGCG..... 21928  
339 tTrpTyrLeuAspAspProValAlaAspProSerLeuVal..... 352  
21927 .....CTGTGGAGCCGCGGTGCTTCCGCGTCTCGCGCCACCG 21885  
353 .....ProLeuTyrPheVal 357  
21884 ACTGCGCGCGCGTACTGGGCGACATGTGCGCTCTACCTGTTC 21835  
358 AlaAlaGluAlaArgLysHisValLysValValLeuSerGlyGluGlyAl 374  
21834 TTCGCGCGGTGCGGAGAACGTACGCTGCGCTCTCGCGGAGCGCG 21785  
374 aspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSerLeu.. 390  
21784 GGACGAACCTCTTCGCGGCTACCGGTGTTCCGCAACCGTTGGCCCTGC 21735  
391 .....AlaProPheGluLysIleProSerProLeuArgLysGlyLeuGly 405  
21734 GCGCGCCACCTTC..... 21721  
406 LysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuG 422  
21720 .....CCGTGGCTCAC 21710  
422 uArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerP 439  
21709 CCGGGTTGC.....GCGGCTACTTCGCGGACGCTCGCTGT 21672  
439 heAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrp 455  
21671 TCAGCGCGGTCTGTGACAGCTGACCTGAGCGGTACCGGGCGGC 21622  
456 AspHisArgGluValThrAlaProIleTyrAlaGlnSerArgAsnPheAs 472  
21621 CGGTACGCGGCGGCTGCGGAAGTGCCTGCTCGCGGTGAGTCGCG 21572  
472 pProValAlaArgMetGln.....HisLeuAspLeuPheThrTrpM 486  
21571 GCCCGAGCGGATCGCGAGATCGGTCACTGAACCTGACCCCGCTTCG 21522  
486 eArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSer 502  
21521 TGCAG...ACCCTGCTCGACCGCAAGCGGATGACGATGCGGTCGCG 21475  
503 LeuGluLeuArgValProPheLeuAspLysGluValPheLysValAlaG 519  
21474 CTCGAGGTGCGGTGCTTCTCGACACCGGCTCGTGAGTACGTCTT 21425  
519 uhrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaL 536  
21424 CAACGTGCTTGGAGTGAAGATTTTCACGCGCGGAGAGAGCGCTC 21375  
536 eArgArgAlaLeuGluGlnIleValProHisValLeuHisArgLys 552  
21374 TCGGCGCGGTGCGGACCTGCTCGCGGTGCTGCTGAGCGGCTG 21325  
553 LysLeuGlyPheProVal.....ProMetAr 561  
21324 AAGACCCGTACCGCGGACCCAGGATCCGGCTACGAGGCGCGCTGCG 21275  
561 gHisTrpLeuAlaGly 566  
21274 GCCCGAAGTGGCGGG 21259  
seq\_name: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-474-933-1  
seq\_documentation\_block:  
; Sequence 1, Application US/08474933

```

; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlortetracycline and tetracycline Forma
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-474-933-1

Ignment_scores:
    Quality: 633.00      Length: 622
    Ratio: 1.829        Gaps: 20
    Percent Similarity: 55.627    Percent Identity: 29.582

alignment_block:
US-09-786-474-2 x US-08-474-933-1/rev ..

Align seg 1/1 to reverse of: US-08-474-933-1 from: 1 to: 30001

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
|||||
22914 ATGTGCGGAATCGTGGGGTGTGCAGTACGACCGCCGCGGCGGACACCG 22865

17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
|||||
22864 CCGGACCTGCTCCACGGCGATGACCGACACCATGCTGCGGGGCGCG 22815

34 sPAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg 50
|||
22814 ACGCGAGGGCGCTCTGTTCCGCGCGCGCGCGGCTCGGTTCACCGCGC 22765

51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67
|||||
22764 CTGTGGGTGATCGACCCGACGCGCGCGCGCGATG.....AC 22724

67 oAlaAspGluProAspArg...TyAlaMetThrPheAsnGlyGluIleT 83

```



```

486 targGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerL 503
   :|||
10131 ACGT .....GCAACAAATCTATGCGACGTTGGGGCA 10100
   :|||

503 euGluLeuArgValProPheLeuAspLysGluValPheLysValAlaGlu 519
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
10099 TCGACGACGCTGTTCTTCTTAGATAAAGAAATTTGTTGATCGCTCAATG 10050
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

520 ThrIleProTyrAspLeuLysIleAlaAsnGly.....ThrThrLysTy 534
   :||| :||| :||| :||| :||| :||| :||| :|||
10049 CGCTAAACCCCTGAACATAAAATGATCACGGCGCGCATTTGAGAAAAA 10000
   :||| :||| :||| :||| :||| :||| :||| :|||

534 rAlaLeuArgArgAlaLeuGluGlnIleValProHisValLeuHisA 551
   :||| :||| :||| :||| :||| :||| :||| :|||
9999 CATCATTCAGAGACGATTTGAAGATTTACTACCTGAAGAAATTTGATGCG 9950
   :||| :||| :||| :||| :||| :||| :||| :|||

551 rgLysLysLeuGlyPhe 556
   :||| :||| :||| :||| :||| :||| :||| :|||
9949 GTCAAAAAGAACATTC 9933
   :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: /cgn2/ptodata/1/ina/backfiles1.seq:5256558-3

seq_documentation_block:
; Patent No. 5256558
; APPLICANT: CORUZZUI, GLORIA M.TSAI, FONG-YING
; TITLE OF INVENTION: GENE ENCODING PLANT ASPARAGINE SYNTHETASE
; NUMBER OF SEQUENCES: 17
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/514,816
; FILING DATE: 26-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 347,302
; FILING DATE: 03-MAY-1989
; SEQ ID NO.:3
; LENGTH: 2002
5256558-3

alignment_scores:
Quality: 425.00 Length: 592
Ratio: 1.384 Gaps: 24
Percent Similarity: 51.858 Percent Identity: 25.845

alignment_block:
US-09-786-474-2 x 5256558-3 ..

Align seg 1/1 to: 5256558-3 from: 1 to: 2002

1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla.. 16
  ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ATGTGTGGTATACTTGTCTGTTGTTGTTCTCATCTCTTCGAGCCAA 151

17 .....PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
   ||| ||||| ||||| ||||| ||||| ||||| |||||
152 GAGAGTTCTGTTGTTGGAATTTTCACGCAGATTG.....AAGCACC 192
   ||||| ||||| ||||| ||||| ||||| |||||

31 rgGlyProAspAlaGlyThrTriphHisAspAlaAspAlaAlaPheGly 47
  ||||| ||||| ||||| ||||| ||||| ||||| |||||
193 GAGCCCTGAATGGAGTGGCTCCACCAATGATGTTATTATTGGCA 242

48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
   :||| :||| :||| :||| :||| :||| :||| :|||
243 CAACAACGCTAGCCATAGTTGATCTCTGTTGATCAACCTCTC.. 290

64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
   :||| :||| :||| :||| :||| :||| :||| :|||
291 .....TTCAATGAAGACACATCGTCAATTTGTCACGGTAAACGGAG 330

81 IuileTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
  ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 AGATTTACATTAATGAAGATCTCAGGAACAGTTGTCTAAT.....CAC 374

98 ThrPheAsnThrSerGlyAspGlyValProIleValValGlyPheHisHI 114

```

Align seq 1/1 to: 5256558-1 from: 1 to: 1886



```
1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla.. 16
|||||
94 ATGTGTCGACATCTGCTGACTTGGTGTCTCTGATGATTCACAGCTAA 143
17 .....PheValProAlaLeuGluAlaLeuProCysMetArgHisA 31
   |||
144 ACGAGTTCGCATCTACGAGCTTCTCCAGATTG.....AAGCAC 184
31 rGlyProAspAlaGlyThrTrpHisAspAlaAspAlaPheGly 47
|||||
185 GTGGCCAGACTGGAGTGGCTCCACCAACATGCTGATACTATTGGCT 234
48 PheAsnArgLeuSerIleAlaPheAlaHisSerHisGlnProLeuAr 64
   |||
235 CATCAAGGTTAGCCATGTTGATCCTGCTGCTGCTGATCACCTC.. 282
64 gTrpGlyProAlaAspGluProAspArgTyAlaMetThrPheAsnGly 81
|||||
283 .....TTCAATGAAGACAAATCAATATTGTCACGGTGAATGGAG 322
81 luIleTyAsnTyValGluLeuArgLysGluLeuSerAspLeuGlyTy 97
|||||
323 AAATCTACAATCATGAAGAGCTCAGAAACAATTGCCTCAAT..... 366
98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHis 114
   |||
367 AGTTTTTTACACATGCTGACTGTGATGTTATTCACACCTGTACGAG 416
114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAl 131
   |||
417 ACATGGAGAAAAATTTCTGGATGTTAGACGGTATATTTCTGTTTGT 466
131 leTrpAspThrLysGlyLysSerLeuPheLeuAlaArgAspGlnPhe 147
   |||
467 TGCTGGATACCTCGACAACAGTTTCATAGTTCGAGGAGTGCCTATAG 516
148 IleLysProLeuPheTyAlaThrGluHisGlyThrValPheSerSe 164
   |||
517 GTTACTCTTGATCATGTTGGGACTAGTGGTCTGTTGG..... 561
164 rGluLysThrIleLeuGluMetAlaGluMetAsnLeuAspLeuG 181
   |||
562 .....ATTGCATCAGAAATTGAAA.....G 580
181 lyLeuAspLysArgThrIleGluHisTyValAspLeuGlnTyValpr 197
   |||
581 GACTGAATGATGAA...TGTAACATTTC.....GAAGTTT 615
198 GluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCys 214
   |||
616 CCGCCCGTCACTTACTCGAGCAAGAAGA..... 648
214 rAlaThrValArgProGlyGlyLysLeuGluGlnLysArgTyThrPhe 231
   |||
649 .....GAGTTTCTCGATGGTATAATC 670
231 roGlnPhe.....ProValGlnLysVal 238
   |||
671 CTCATGGTTCATGAGGCTAATATTCGTCACACCTTATGATCCTCTA 720
239 ValLysGlyLysGluGlnAspLeuPheAspArgIleAlaGlnValLeu 255
   |||
721 GTT.....TTGAGAACGCGTTGA 740
255 uAspSerValGluLysHisMetArgAlaAspValThrValGlySerPhe 272
   |||
741 GAAGGCTGTGATAAAGAGTTGATGACCGGATGTCCTTTCGGGGTTTAC 790
272 euSerGlyGlyIleAspSerThrAlaIleAlaProLeu..... 284
   |||
791 TATCGGAGGTTTGGATTTCATCGTGGTGGCGTCTGTCACCTGCTAGATAC 840
```

seq\_name: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-737-825-1



```
seq_documentation_block:
; Sequence 16, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A. Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jennifer Buliding, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-385-028-16

alignment_scores:
Quality: 165.00 Length: 567
Ratio: 0.693 Gaps: 23
Percent Similarity: 41.975 Percent Identity: 21.869

alignment_block:
US-09-786-474-2 x US-09-385-028-16 ..
Align seg 1/1 to: US-09-385-028-16 from: 1 to: 1542

18 ValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyPro.. 33
|||||:|||||:
33 GTTCTGGCTCCGCGC.....CCGAACGGGGGGGGGGGGGGGGGGCG 73
34 .....AspAspAlaGlyThrTrpHis.AspAlaAspAlaAlaPhe 46
|||||:|||||:
74 GCCCGGCTTGGCCCGGGGGAGCCACACCGACATCGACACGCCCGCAG 123
47 GlyPheAsnArgLeuSerIleAlaHisSerHisGlnProLe 63
|||||:|||||:
124 GGGAGGGCTGCTCGCG.....GCCACCTGGTGACAGCC..... 159
63 uArgTrpGlyProAlaAspGluProAspArg..... 73
|||||:|||||:
160 .....CCCTCGGTGGCGCGCGCGGGTGGCGGCTCCCTCA 199
```

```
74 .....TyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrVal 86
200 CCGGGCGCCACACCGCGGTGCTCGCGGTGAGATCTACAACCGGGAC 249
87 GluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerG 103
|||||:|||||:
250 GAACCTC.....CTCTCGTGTGCTCGCGCGGACCCCGCGCGAGGG 290
103 yAspGlyGluProIleValValGlyPheHisHisTrpGlyGluSerVal 120
|||||:|||||:
291 GGACGGGAGCTGCTCGCGGTGCTGGAACGCTATGACCTGATGCTGCT 340
120 alGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLys 136
|||||:|||||:
341 TCCGGCTGGTGAACGGCGCTTCGGACCGGTGGT...CGGACCGGGAC 387
137 LysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysProLeuPhe 153
|||||:|||||:
388 CGGTCTCTG...CTGCCACCGACCGCGGTTCGGTCCGCTGTACAC 434
153 rAlaThrThrGluHisGlyThrValPheSerSerGluLysLysThrIle 170
|||||:|||||:
435 CTGTGTGGCGCGCGGAGGTCCGGGCTCCACCGAGGCCAAGCGCTCG 484
170 euGluMetAlaGluMetAsnLeuAspLeuGlyLeuAspLysArgThr 186
|||||:|||||:
485 CCGCGCACCGCGCGCGGAGGCTTCGCGCTCGCG...GACCGCGCGGG 531
187 IleGluHisTyrValAspLeuGlnTyrValProGluProAspThrLeu 203
|||||:|||||:
532 CTGCGCGTCTGACCGGTGCTACCGAGTGCCTCGCGCGCGCGCTGATGA 581
203 sAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValArg.... 218
|||||:|||||:
582 CATCGAC.....CTCGGCTCGGCGACCGCGCTCACCCACCGCACCT 622
219 .....ProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPhe 233
|||||:|||||:
623 GGACCGCGGGC.....CTCTCCCGCGCATCTCGCGGAGGGC 660
234 ProValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArg 250
|||||:|||||:
661 GAGCGCGTGGCGCGTGTGCGGCGC..... 684
250 eAlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspVal 267
|||||:|||||:
685 .....GGCTGGAGAGGCGCTCGCGCGAGCGGTCA 715
267 hrValGlySer.....PheLeuSerGlyGlyIleAspSerThr 279
|||||:|||||:
716 CCGCGCGGACACCGCGGTGTGTGCTCTCGCGGGAATCGACTCTCC 765
280 AlaIleAlaProLeuAlaLysArgHisAsnProAspLeuThrPheTh 296
|||||:|||||:
766 GGGTGGCGGCTGTGCGCACCGCGCGCGGGAACCTGGACACCGGTGTC 815
296 rThrGlyPheGlu.....ArgGluGlyTyrSerGluValA 308
|||||:|||||:
816 CATGGGACCGACACGCTCCAACGAGTTCGCGGAGCGCGCGGTGCTCG 865
308 spValAlaAlaGluSerAlaAlaIleGlyAlaGluHisIleValLys 324
|||||:|||||:
866 AC.....CATCTGGCGACCGCGCGCGGAGATC 894
325 IleValSerProGluGluTyrAlaAsnAlaIleProLysIleMetTrp 341
|||||:|||||:
895 ACCATCCGACCGAGCTGCTGGCGAGCTCCCGTACGGGTGTGGGC 944
341 rLeuAsp.....AspProValAlaAspProSerLeuValProLeuTyr 356
|||||:|||||:
945 CTCCGAGTGTGGACCGGACATCATCGAGTACCTGCTCCCTCCGACAG 994
```

```
356 heValAlaAlaGluAlaArgLysHisValLysValValLeuSerGlyGlu 372
   :::: :::::|||||
995 CGCTCTACGGGGCGCTCGACGGCGCGGCGCATCTCCACGGGTAC 1044
   :::::|||||
373 GlyAlaAspGluLeuPheGlyGlyThrIleTyrLysGluProLeuSe 389
   |||||:::|||||
1045 GCGCGGACATCCCTCTCGGGGC.....ATGCACCGGAG..... 1080
   :::::|||||
389 rLeuAlaProPheGluLysIleProSerProLeuArgLysGlyLeuGlyL 406
   :::::|||||
1081 .....GACCGGTGCCCGG..... 1095
   :::::|||||
406 ysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGlu 422
   |||||:::|||||
1096 ..CTGGACACCGTTCTCGCGCAGCATG..... 1122
   :::::|||||
423 ArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerPh 439
   |||||
1123 .....GCCACCTTGA 1133
   :::::|||||
439 eAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrpA 456
   :::::|||||
1134 CGGGCTGAACGAGATCTCCCGTGTGTCCACGCTGGCGGGCAGCTGA 1183
   :::::|||||
456 sPHisArgGluValThrAlaProIleTyrAlaGlnSerArgAsnPheAsp 472
   :::::|||||
1184 CCACCCAC..... 1191
   :::::|||||
473 ProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAs 489
   :::::|||||
1191 ..... 1191
   :::::|||||
489 pileLeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuA 506
   :::::|||||
1191 ..... 1191
   :::::|||||
506 rgValProPheLeuAspLysGluValPheLysValAlaGluThrIlePro 522
   |||||:::|||||
1192 ....CCGTACTGGGAGCGGGAGTCTCGATCTGTGTGTGTGTGTGTGTGT 1236
   :::::|||||
523 TyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaLeuArgArgAl 539
   |||||:::|||||
1237 GCGGGCTCAAGCGGGCGGACGGCGGGGACAAGTGGTGTGTGTGTGTGTGT 1286
   :::::|||||
539 aLeuGluGlnIleValProHisValLeuHisArgLysLysLeuGly 555
   |||||:::|||||
1287 GATGGCGGAGCGCTCCCGCGGAGACCGTCAACCGGCCAAGCTGGGC 1335
   :::::|||||
```

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:US-09-385-028-13

```
documentation_block:
  quence 13, Application US/09385028
  Patent No. 6232106
  GENERAL INFORMATION:
  APPLICANT: Susan E. Jensen
  APPLICANT: Kwamena A. Aldoo
  APPLICANT: Ashish S. Paradkar
  TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
  Patent No. 6232106
  TITLE OF INVENTION: Acid Biosynthesis
  NUMBER OF SEQUENCES: 25
  CORRESPONDENCE ADDRESS:
  ADDRESS: JACOBSON, PRICE, HOLMAN & STERN, PPLC
  STREET: The Jenner Building, 400 Seventh Street, N.W.
  CITY: Washington
  STATE: D.C.
  COUNTRY: U.S.A.
  ZIP: 20004
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-385-028-13

alignment_scores:
  Quality: 165.00      Length: 567
  Ratio: 0.693        Gaps: 23
  Percent Similarity: 41.975  Percent Identity: 21.869

alignment_block:
  US-09-786-474-2 x US-09-385-028-13 ..
  Align seg 1/1 to: US-09-385-028-13 from: 1 to: 11604

18 ValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyPro.. 33
   |||||:::|||||
1940 GTTCTCTGGCTCCGC.....CGAACGGGGGGGGGGGGGGGGGG 1980
   :::::|||||
34 .....AspAspAlaGlyThrTrpHis AspAlaAspAlaAlaPhe 46
   |||||:::|||||
1981 GCGCGGTCTTCGGGACCGCGGCGGACGACATCGACAGCCCGCCAG 2030
   :::::|||||
47 GlyPheAsnArgLeuSerIleLeuAspIleAlaHisSerHisGlnPro 63
   |||||:::|||||
2031 GGGGAGCGCTCGCTCGG...GCGACCTGTGTGCACGCC..... 2066
   :::::|||||
63 uArgTrpGlyProAlaAspGluProAspArg..... 73
   |||||:::|||||
2067 .....CCTCGGTGCGCGCGGACCGCGGCGGTGCGCGCTCCCTCA 2106
   :::::|||||
74 .....TyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrVal 86
   |||||:::|||||
2107 CCGGCGCGCGCGCGCGCGGTGCTCGCGGTGAGATCTACAACCGGGAC 2156
   :::::|||||
87 GluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerG1 103
   |||||:::|||||
2157 GAATCT.....CTCTCCGTGTGTGCGCGCGGACCGCGCGGAGGG 2197
   :::::|||||
103 yAspGlyGluProIleValIleValGlyPheHisTrpGlyGluSerValV 120
   |||||:::|||||
2198 GAGCGGAGAGTGTCTCGGCTGCTGGAACGCTATGACCTGCATGCGCT 2247
   :::::|||||
120 aIcLHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGlu 136
   :::::|||||
2248 TCGGCTGTGAACGGCGCTTCGCGACCGTGGTG...CGGACCGGGGAC 2294
   :::::|||||
137 LysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysProLeuPheTy 153
   :::::|||||
2295 CGGGTCTG...CTGCCACCGACCGCGGTTCGGTCCGCTGTATAC 2341
   :::::|||||
153 rAlaThrThrGluHisGlyThrValPheSerSerGluLysLysThrIleL 170
   :::::|||||
```

2342 CTGTGTCGGCGCGGAGTCTCCGGGGTGCTCACGAGGCCAACAGCGCGTCG 2399

170 euGlMetAlaGluMetAsnLeuAspLeuGlyLeuAspLysArgThr 186  
::: ::::: |||::: |||  
2392 CCGCACCACCGGCACCCGAAGGCTTCCCGTCGG...GACGCCGCGCG 2438

187 lleGluHisTrpValAspLeuGlnTyrrValProGluProAspThrLeuH1 203  
::: ::::: ||| |  
2439 CTCGCGCGTCTGACCGGTGTCTACAGGTGCCCGCGGCGCGGTATGGA 2488

203 saLaGlnIleSerArgLeuGluSerGlyCystrAlaThrValArg.... 218  
2489 CATCGAC.....CTCGCTCGGCACCGCGGTCAACCCACCGCACCT 2529

219 ....ProGlyGlyLysLeuGlnLysArgTyrrPheLysProGlnPhe 233  
||| |::: |||::: |||:::  
2530 GGACCCCGGC.....CTCTCCCGCCCATCTCTCCGGAGGC 2567

234 ProValGlnLysValLysValLysGlyGluGlnAspLeuPheAspArgII 250  
::: ::::: |||:::  
2568 GAGCGCGTGGCGCGTGGCGGC.....GGCTGGAGAAGCGGTGCCACCGGCTCA 2622

250 eaLaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValt 267  
::: |||::: |||::: |||::: |||:::  
267 hrValGlySer.....PheLeuSerGlyGlyIleAspSerThr 279  
|| |||::: ||| |::: ||| |::: ||| |::: ||| |:::  
2623 CCCCGGGGACACCCGTTGGTGTCTCTCCGCGGAATCGACTCTCTCC 2672

280 AlaIleLeuProLeuAlaLysArgHisAsnProAspLeuLeuThrPheTh 296  
::: ||| |::: ||| |::: ||| |::: ||| |:::  
2673 GSGTCGCGGCTGTGCGCACCGGCGGGAACTGCACACGGGTG 2722

296 rThrGlyPheGlu.....ArgGluGlyTyrrSerGluValA 308  
::: |||::: ||| |::: ||| |::: ||| |:::  
2723 CATGGCACCGACACAGCTCCACGATTCTCCGCGAGGCCGGCGGTCTCG 2772

308 spValAlaAlaGluSerAlaAlaIleGlyAlaGluHisIleValLys 324  
2773 AC.....CATCTGGCCACCCCGCACCGGGAGATC 2801

325 ileValSerProGluGluTyrrAlaAsnAlaIleProLysIleMetTrpy 341  
::: |||::: ||| |::: ||| |::: ||| |:::  
2802 ACCATCCCGACACCGAGCTCTGGCGAGTCCCGTACGCGGTGTGGC 2851

341 leuAsp.....AspProValAlaAspProSerLeuValProLeuTyrrP 356  
::: ||| |::: ||| |::: ||| |:::  
2852 CTCGCGAGTCGTGGACCGCGACATCATCGTAGTACCTCTCCCCGTGAC 2901

356 heValAlaAlaGluAlaArgLysHisValLysValLeuSerGlyGlu 372  
::: |||::: ||| |::: ||| |::: ||| |:::  
2902 CGCTCTACGGGCGTCTCACGGCGGAGCCCGCATCTTCACCGGGTAC 2951

373 GlyAlaAspGluLeuPheGlyGlyTyrrThrIleTyrrLysGluProLeuSe 389  
||| |::: ||| |::: ||| |::: ||| |:::  
2952 GCGCGGACATCCCTCTCGGGC.....ATGACCCCGAG... 2987

389 rleuAlaProPheGluLysIleProSerProLeuArgLysGlyLeuGlyL 406  
.....GACCGGTCTCCCGG..... 3002

406 ysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGlu 422  
||| |::: ||| |::: ||| |:::  
3003 ...CTGGACACCGTCTCCGGCAGCATG..... 3029

423 ArgGlySerMetThrMetGluGluArgTyrrGlyAsnAlaArgSerPh 439  
|||  
3030 .....GCCACTTCGA 3040

439 eaSnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrpA 456  
::: ||| |::: ||| |::: ||| |:::  
3041 CGCGCTGAACAGAGATGCCCGCGTGTCTCCACGCTCGCGGGGACATGGA 3090

```

456  sphisargluValThrAlaProIleTyrAlaGlnSerArgAsnPheAsp 472
      ::::
3091  CCACCCAC..... 3098
      ::::
473  ProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAs 489
3098  ..... 3098
      ::::
489  pileLeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuA 506
3098  ..... 3098
      ::::
506  rgValProPheLeuAspLysGluValPheLysValAlaGluThrIlePro 522
      ::::
3099  ....CCGACTGGGACCGGGAGGTCCTCGATCTGCTGGTCTCGCTGGAG 3143
      ::::
523  TyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaLeuArgArgAl 539
      ::::
3144  GCGGGCTCAAGCGGGCGGACGCGCGGACAACTGGTGTGCGCGCGCGC 3193
      ::::
539  aLeuGluGlnIleValProProHisValLeuHisArgLysLeuGly 555
      ::::
3194  GATGCGCGACGCGCCCTCCGCGGAGACAGCGTCAACCGGCCCAAGCTGGC 3242
      ::::
seq_name: /cgn2_6/ptodata/1/1na/6B_COMB.seq:US-09-385-028-1
seq_documentation_block:
; Sequence 1, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavu
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN,PPLC
; STREET: The Jenifer Buliding, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15079 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

```

ORIGINAL SOURCE:  
ORGANISM: Streptomyces clavuligerus  
US-09-385-028-1

## alignment\_scores:

Quality: 165.00 Length: 567  
Ratio: 0.693 Gaps: 23  
Percent Similarity: 41.975 Percent Identity: 21.869

## alignment\_block:

US-09-786-474-2 x US-09-385-028-1 ..

Align seg 1/1 to: US-09-385-028-1 from: 1 to: 15079

```

18 ValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyPro... 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3972 GTTCCTGGCTCCGC.....CCGACGGCGGGGGGGGGGGGGGGGG 4012
34 .....AspAspAlaGlyThrTrpHis..AspAlaAspAlaAlaPhe 46
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4013 GCCCGGTCTTCGGACCCGGGGGAGGCACACACGACATCGACACGCCCCAG 4062
47 GlyPheAsnArgLeuSerIleIleAspIleAlaHisSerHisInProle 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4063 GGGGAGCGCTCGCTCGCG...GCGACCTGGTGCACGCC..... 4098
63 uArgTrpGlyProAlaAspGluProAspArg..... 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4099 .....CCTCGGTCCGCGCCGACCGCGGGTGGCGGCTCCCTCA 4138
74 .....TyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrVal 86
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4139 CCGCGCGCCACACCGCGGTGCTCGCGGTGAGATCTACACCGGGAC 4188
87 GluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGI 103
||||| |||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4189 GAATC.....CTCTCCGTGCTGCCGCGCGCGCGCGCGGAGGG 4229
103 YAspGlyGluProIleValValGlyPheHisHisTrpGlyGluSerVal 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4230 GGACGGGAGCTGTCTCGGGTGTGTGGNACCTATGACCTCATGCCT 4279
120 alGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGlu 136
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4280 TCCGGGTGTGAACGGCGGCTCGCGACCGTGGTG...CGGACCGGGAC 4326
137 LysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysProLeuPheTy 153
::: ||| |||||| |||: ||| ||| ||| ||| ||| ||| ||| |||
4327 CGGGTCCTG...CTCGCCACCGACCGCGGGTTCGGTGGCGGTATACAC 4373
153 rAlaThrThrGluHisGlyThrValPheSerSerGluLysLysThrIleL 170
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4374 CTGTGTGGCCCGGGCGAGTCCGGCGTCCACCGAGGCCAAGCGCTCG 4423
170 euGluMetAlaGluMetAsnLeuAspLeuGlyLeuAspLysArgThr 186
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4424 CCGCGCACCGCGACCGAAGGGTTCGCGCTCGCG...GACGCCCGCGG 4470
187 IleGluHisTyrValAspLeuGlnTyrValProGluProAspThrLeuHI 203
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4471 GTCGCGGCTGACCGGTGCTACCAAGGTGCCCGGGCGCGCGGTATGGA 4520
203 sAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValArg.... 218
::: ||| |||||| |||: |||: |||: |||: |||: |||: |||: |||
4521 CATCGAC.....CTCGGTCCGGGACCGCGCGGTACCCACCGCACCT 4561
219 .....ProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPhe 233
||||| |||||| |||: |||: |||: |||: |||: |||: |||: |||
4562 GGACCCCGGGC.....CTCTCCCGCGCATCTCTGCGCGGAGGC 4599
234 ProValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgII 250
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||

```

```

4600 GAGGCGCTCGCGCGCGCTCGCGGCC..... 4623
250 eAlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValT 267
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4624 .....GCGCTGGAGAGAGCGCTCGCGCGCGCGGTCA 4654
267 hrValGlySer.....PheLeuSerGlyGlyIleAspSerThr 279
|| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4655 CCCCCGCGACACCCCGTGTGGTGTCTCTCCGCGGAATCGACTCCTCC 4704
280 AlaIleAlaProLeuAlaLysArgHisAsnProAspLeuLeuThrPheTh 296
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4705 GGGGTGCGCGCTGTGCGCACCGCGCGCGCGGGAACATGGACACGCTGC 4754
296 rThrGlyPheGlu.....ArgGluGlyTyrSerGluValA 308
: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4755 CATGGGACCGACACGTCACACGAGTTCGCGAGCGCGCGCGGTGCTGC 4804
308 spValAlaAlaGluSerAlaAlaIleGlyAlaGluHisIleValLys 324
|| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4805 AC.....CATCTGCGCACCGCGCACCGCGGAGATC 4833
325 IleValSerProGluTyrAlaAsnAlaIleProLysIleMetTrpTy 341
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4834 ACCATCCGACACCGAGCTGTGGCGCAGCTCCGTCACGCGGTGGGC 4883
341 rLeuAsp.....AspProValAlaAspProSerLeuValProLeuTyrP 356
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4884 CTCCGAGTGGTGGACCGCGACATCATCGAGTACCTGCTCCCTCGACAG 4933
356 heValAlaAlaGluAlaArgLysHisValLysValValLeuSerGlyGlu 372
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4934 CGCTCTACCGGGCGCTCGACGGCGGAGCGCGCATCTCTACCGGGTAC 4983
373 GlyAlaAspLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSe 389
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
4984 GCGCGGACATCCCTCGGGGC.....ATGACCGCGGAG..... 5019
389 rLeuAlaProPheGluLysIleProSerProLeuArgLysGlyLeuGlyL 406
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
5020 .....GACCGGTGCGCGCG..... 5034
406 ysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGlu 422
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
5035 ..CTGGACACCGTCTCGCGCACGACATG..... 5061
423 ArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerPh 439
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5062 .....GCCACCTCGA 5072
439 eAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrpA 456
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
5073 CGGGCTGAACGAGATGTCCTCGGTGTCTCCACGCTGGCGGGCGCACTGGA 5122
456 spHisArgGluValThrAlaProIleTyrAlaGlnSerArgAsnPheAsp 472
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
5123 CCACCCAC..... 5130
473 ProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAs 489
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5130 ..... 5130
489 pIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuA 506
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5130 ..... 5130
506 rgValProPheLeuAspLysGluValPheLysValAlaGluThrIlePro 522
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
5131 .....CCGTACTGGACCGGGAGGTCTCGATCTGCTGTGCTCTCGCTGGAG 5175
523 TyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaLeuArgAl 539
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
5176 GCGGGCTCAAGCGCGCGCACCGCGCGGACAGTGGGTGCTCGCGCGCGC 5225

```

```

539 aLeuGluClnIleValProProHisValLeuHisArgLysLysLeuGly 555
      |:::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
5226 GATGGCGAGCCCTCCGGGGAGACGCTCAACGGCCCAAGCTGGC 5274
seq_name: /cqn2.6/ptodata/1/lna/5A_COMB.seq:US-08-599-171A-22

```

seq\_documentation\_block:  
; Sequence 22, Application US/08599171A  
; Patent No. 5814473

: GENERAL INFORMATION:  
 : APPLICANT: WARREN, Patrick V.  
 : TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
 : NUMBER OF SEQUENCES: 32  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
 : ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 : STREET: 6 BECKER FARM ROAD  
 : CITY: ROSELAND  
 : STATE: NEW JERSEY  
 : COUNTRY: USA

```

: ZIP: 07068
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: 3.5 INCH DISKETTE
:
: COMPUTER: IBM PS/2
:
: OPERATING SYSTEM: MS-DOS
:
: SOFTWARE: WORD PERFECT 5.1
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/599.171A
:
: FILING DATE: Concurrently
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER:
:

```

```

: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: HERRON, CHARLES J.
: REGISTRATION NUMBER: 28,019
: REFERENCE/DOCKET NUMBER: 331400-38
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1779 NUCLEOTIDES
: TYPE: NUCLEIC ACID
: STRANDEDNESS: SINGLE
: TOPOLOGY: LINEAR
: MOLECULE TYPE: GENOMIC DNA
: US-08-599-171A-22

```

```
alignment_scores:
  Quality: 154.50
  Ratio: 0.592
  Percent Similarity: 45.000
  Percent Identity: 20.000
  Length: 580
  Gaps: 27
```

```
alignment_block:
US-09-786-474-2 x US-08-599-171A-22
..
```

Align seq 1/1 to: US-08-599-171A-22 from: 1 to: 1779

1 MetCysGlyLeuLeuGlyTlleLeuThrAlaAsnAlaGluAlaph 17  
 1 ATGTCGGGGTATGTCGGATACGTA.....GGGAGGGATTAGC 38  
 17 evalProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34  
 39 CTTTCCTATGCTCCGGAGCTCTTGAGACATCGAATACAGGGGTAC 88

34 spAspAlaGlyThr..... 38  
 ||:::|||||:::  
 89 ACTCCGGGGGAGTTTGCCTTTATAGAAGACGGGAACCTCATAGTTGAAAAG 138

39 .....Trrp.....HI 40  
139 AAGAGGGAAATAAGGGAACCTCGTTAAAGCGCTATGGGAAAGGATTA 188  
40 saspAlaAspAlaPheGlyPheAsnArgLeuSer..... 52  
189 CAAGGCTAAACACGGGTATAGGTCACACACGCTGGGCAACCCACGGAAAGC 238  
53 ..lleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla 68  
239 CCAGGACGAGACGCCACCCAC.....ACC 267  
69 AspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyr 85  
268 GACCAAAAGGTGAGTTTGCACTAGTTTACACACGGGATTAAGAAACATA 317  
85 rValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThr 102  
318 CTTAGAACTAAAGAGAACTAAGAGGAAGGTGTAAGATTTCAGGTCCG 367  
102 erGlyAspGlyGluProIle.....ValValGlyPheHisHisTrpGly 116  
368 AAACAGACACAGAAAGTTATAGCCCACTCATAGGAAGAACAATACAGGGG 417  
117 Glu.....SerValValGluHisLeuArgGlyMetPhe 127  
418 GACTTACTGGAGGCGGTTTAAACACCGTAAGAANAATTAAGGGTGCTTT 467  
127 eGlyIleAlaIleTrpAspThrLysGluLysSerLeuPheLeuAlaArg 144  
468 TGCCTTTGCGTTATAACGGTTTACGAACCAACAGACTAATA..... 510  
144 spGlnPheGlyIleLys.....ProLeuPheTyrAlaThrThrGlu 157  
511 .....GAGTGAACACGGGGAGTCTCTTAATCGTCGGACTCGGAGAA 552  
158 HisGlyThrValPheSerSerGluLysLysThrIleLeuGluMetAlaGln 174  
553 GGAGAAACCTCCTCGCTTCAGATATCCCGCAATACTTCTTACACGAA 602  
174 uGluMet..... 176  
603 AAAGATTATTGTTCTTGATGACGGGGAAATACGGACCTGACTCCCGACA 652  
177 .....AsnLeuAsp..... 179  
653 CTGTGAACATTACACTTTGGGGAGAGCCGTTTCAAAGGAAGTAATG 702  
180 .....LeuGlyLeuAspLysArgThrIleGluHisTyr 190  
703 ATTACGCCCTGGGATCTGTTTCTCGGAAAAGGTGTTTTAAACACTT 752  
190 rValAspLeuGlnTyrValProGluPro.....AspThrLeu 203  
753 CATGCTAAAGAGATATACGAACAGCCCAAGCCATAAACGACACACTCA 802  
203 isAlaGlnIleSerArgLeuGluSer..... 211  
803 AGGTTTTCTCTCAACCGAAGCAATACCTTTTAAGTTAAAGACTTC 852  
212 .....GlyCysThrAlaThrValArgProGly..... 220  
853 AGAAGGCTTTTAATATAGCTGCGGGACCTCTTACCAACGGCGGCTCGT 902  
221 .GlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnL 237  
903 CGGAAAGTACTGGATAGAGATT.....GCAGGTGTTCCACAGAGG 946  
237 ysVal.....ValLysGlyLysGluGln 244  
947 TAATTTACGCTTCGGAAATTCAGGTATCGGACGTTCCGCTTTCGGACAA 996

```
245 AspLeuPheAspArgIleAlaGln..... 252
||||: |||:||||
997 GATATCGTTATTCGAATTTCCAGTCCAGGAGACCGCTGACACAAAGTT 1046
253 ValLeuGluAspSerValGluLysHisMetArgAlaAspValThrValG 269
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
1047 TGGCCTTCAGTCGCAAGAAAG.....GAGCCCTTACCGTGG 1087
269 ly.....SerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaPro 283
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1088 GACTCGTAAACGTAGTGGGAAGTGCCATAGACACAGGAGTCGGACTTTCC 1137
284 LeuAlaLysArgHisAsnProAspLeu.....LeuThrPh 295
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1138 CTTACACACATCGCGACCGACCGAATAGCGGTGGCGGTACAAAGACCTT 1187
295 eThrThrGlyPhe.....G 300
||||: |||
1188 CACCGCACATTCACCGCACTCTACCGCCCTTCGCTAAGGGAAGTGAGG 1237
300 luArgGluGly.....TyrSerGluValAlaAspValAlaAlaGlu 312
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1238 AGAGGGAATCTAATAAGACTCTTGGAAAGGTTCCATCAGCTGTTGAA 1287
313 SerAlaAlaIleGlyAlaGluHisIleValLysIleValSerProGl 329
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1288 CAAACACTGAACACC...GCAGAAGAAGTGGAGAAGGTAGCGGAAAGTA 1334
329 uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProv 346
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1335 CATGAAAGAAAGAAACATGCTTTACCTCGGAAGGTACTTAATTTACCCCA 1384
346 aAla.....AspProSerLeuValProLeuTyrPheValAlaAla 359
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1385 TAGCGCTGGAGGAGCTCTTAACCTTAAGAAATTTCTACATACACGCG 1434
360 GluAla.....ArgLysHisValLysValValLysSerGl 371
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1435 GAAGGTATCCCGACGGGAGATGAAGCAGCGTCCCATAGCCCTCATAGA 1484
371 yGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProL 388
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1485 CGAAACATCCCGTGTGGTA..... 1506
388 euSerLeuAlaPro.....PheGluLysIleProSerProLeu 400
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1507 ....ATCGCACCGAAGACAGCGGTTTACGAGAAGATACTCTCAAACGTA 1551
401 ArgLysGlyLeuLysLysSerLysValLysValProAspGlyMetLysGl 417
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1552 GAAGAGGTTCTCGCAAGAAGGAAGGTTATTCTGTAGGCTTTAAGG 1601
417 yLysSerLeuLysGluArgGlySerMetMetGluGlu 430
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
1602 AGACGAAACTCTCAAAGCAATCCGAGAGCGGTATGGAA 1641
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-646-590B-22
seq_documentation_block:
; Sequence 22, Application US/08646590B
; Patent No. 5962283
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646.590B
; FILING DATE: 08-May-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01094
; FILING DATE: 21-January-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1779 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1776
US-08-646-590B-22

alignment_scores:
Quality: 154.50 Length: 580
Ratio: 0.592 Gaps: 27
Percent Similarity: 45.000 Percent Identity: 20.000

alignment_block:
US-09-786-474-2 x US-08-646-590B-22 ..

Align seg 1/1 to: US-08-646-590B-22 from: 1 to: 1779
1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
1 ATGTGCGGATAGTCGGATACGTA.....GGGAGGATTTAGC 38
17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
39 CCTTCCTATAGTCTCGGAGCTCTTGAGAGACTCGAATACAGGGGTACG 88
34 spAspAlaGlyThr..... 38
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||
89 ACTCCGCGGAGTTGCCCTTATAGAACACGGGAAACTCATAGTTGAAAG 138
39 .....Tip.....H1 40
139 AGAAGCGGAAGATAAGGAACTCGTTAAGCCCTATGGGAAAGGATTA 188
40 sAspAlaAspAlaAlaPheGlyPheAsnArgLeuSer..... 52
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
189 CAAGGCTAAACGGGTATAGGTACACACGCTGGGCAACCCACGGAAGC 238
53 ..IleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla 68
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
239 CCACGGACGAGAACGCCACCCCCAC.....ACC 267
69 AspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTy 85
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
268 GACGAAAAAGGTCAGTTTCAGTAGTTCACACGGGATATAGAAAACTA 317
```



85 rValGluLeuArgLysGluLeuSerAspLeuGlyThrPheAsnThrS 102  
318 CTTAGAACTAAAAGAGAACTAAAGAAAGAAAGTTTAAAGTTCAGGTCCG 367  
102 erGlyAspGlyClnuProfile.....ValValGlyPheHisHisTrpGly 116  
368 AAACAGACACAGAGATTATAGCCCACTCATAGCGAAGAACTACAGGGGG 417  
117 Glu.....SerValValGluHisLeuArgGlyMetPh 127  
418 GACTTACTGGAGCGCTTTAAAAAACCGTAAGAAATTAAGAGGTGCTTT 467  
127 egGlyLeuAlaIleTrpAspThrLysGluLysSerLeuPheLeuAlaArgA 144  
468 TGCCCTTCGGGTATTAACGGTTCACGAACCAACAGACTAATA..... 510  
144 spGlnPheGlyIleLys.....ProLeuPheTrAlaThrThrGlu 157  
511 .....GGAGTGAAGCAGGGAGTCTTTAAATCGTCGGACTCGGAGAA 552  
158 HisGlyThrValPheSerSerGluLysLysThrIleLeuGluMetAlaG 174  
553 GGAGAAAACTCTCTCGCTTCAGATATTCCCGCAATACTTCTCTACACGAA 602  
174 uGluMet.....ValLysGlyLysGluGln 176  
603 AAGATTATTGTTCTGTATGATGCGGGGAATAAGCGGACTGACTCCCGACA 652  
177 .....AsnLeuAsp..... 179  
653 CTCTGAACATTTTACAACCTTTGAGGAGAGCCGCTTCAAAGGAAGTAATG 702  
180 .....LeuGlyLeuAspLysArgThrIleGluHisTy 190  
703 ATTACGCCCTGGGATCTGTTCTTCGGAAAGGGTGTTTTAAACACTT 752  
190 rValAspLeuGlnTrpValProGluPro.....AspThrLeuH 203  
753 CATGCTAAAAGAGATATACGAACAGCCCAAGCATAAACGACACACATCA 802  
203 isAlaGlnIleSerArgLeuGluSer..... 211  
803 AGGGTTCTCTCAACCGAAGACGCAATACCCTTTAAGTTAAAGACTTC 852  
212 .....GlyCysThrAlaThrValArgProGly..... 220  
853 AGAAGGTTTTAATAATACGTGCGGGACCTCTTACCACGGCGGCTTCGT 902  
221 .GlyLysLeuGluGlnLysArgThrPheLysProGlnPheProValGlnL 237  
903 CGGAAAGTACTGGATAGAGATTTT.....GCAGGTGTTCCCAACAGAGG 946  
237 ysVal.....ValLysGlyLysGluGln 244  
947 TAAATTTACGCTTCGGAATTCAGGTATGCGGAGTTCGGTTCGGACAG 996  
245 AspLeuPheAspArgIleAlaGln..... 252  
997 GATATCGTTATCGGAATTTCCCAAGTTCGAGGACGCGTGACACAAAAGTT 1046  
253 .ValLeuAspSerValGluLysHisMetArgAlaAspValThrValG 269  
1047 TGCCCTTCAGTCCGCAAGGAAAG.....GAGACCTTTACCGTGG 1087  
269 ly.....SerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaPro 283  
1088 GACTCGTAAACCTAGTGGGAAGTCCATAGACGGAGTCGGACTTTTCC 1137  
284 LeuAlaLysArgHisAsnProAspLeu.....LeuThrPh 295  
1138 CTTTACACATCGCGGACCGGAATAGCGTGGCGGCTTCAAGACCTT 1187

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1779 NUCLEOTIDES  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: SINGLE  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE: GENOMIC DNA  
 US-09-069-226-22

## alignment\_scores:

Quality: 154.50 Length: 580  
 Ratio: 0.592 Gaps: 27  
 Percent Similarity: 45.000 Percent Identity: 20.000

## alignment\_block:

US-09-786-474-2 x US-09-069-226-22 ..

ign seg 1/1 to: US-09-069-226-22 from: 1 to: 1779

```

1 MetCysGlyLeuLeuGlyLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
  |||||
1 ATCTGCGGATAGTCGGATACGTA.....GGGAGGATTTAGC 38

17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
  |||||
39 CCTTCCTATAGTCTCGGAGCTCTTGAGAGACTCGAATACAGGGGTACG 88

34 sAspAlaGlyThr..... 38
  |||||
89 ACTCCGGGGAGTTGCCCTTATAGACAGCGGAACTCATAGTTGAAG 138

39 .....Trp.....H1 40
  |||

139 AAGAGGGAAGATAAGGAACTCGTTAAAGCGCTATGGGAAAGATTAA 188

40 sAspAlaAspAlaPheGlyPheAsnArgLeuSer..... 52
  |||

189 CAAGGCTAAACGGGTATAGTGCACACCTCGTGGCAACCCAGCGAAAGC 238

53 ..lleleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla 68
  |||||
239 CCACGGACGAGACGCCACCCAC.....ACC 267

69 AspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTy 85
  |||||
268 GACGAAAGGTCAGTTGCAGTAGTTCACACCGGGATATAGAAACTA 317

85 rValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrs 102
  |||||
318 CTTAGAACTAAAGAGAACTAAAGAGGAGGTGTAAAGTTCAGGTCCG 367

102 erGlyAspGlyGluProIle.....ValValGlyPheHisHisTrpGly 116
  |||||
368 AAACAGACACAGAAAGTTATAGCCCACTCATACGGAAGAACTACAGGGG 417

117 Glu.....SerValValGluHisLeuArgGlyMetPh 127
  |||||
418 GACTTACTGGAGCCGTTTAAACCGTAAAGAAATTAAGGGTGTCTTT 467

127 eGlyIleAlaIleTrpAspThrLysGluLysSerLeuPheLeuAlaArgA 144
  |||||
468 TGCTTTGGGTTATACGGTTCCACCAACCAACACACTAATA..... 510

144 spGlnPheGlyIleLys.....ProLeuPheTyrAlaThrGlu 157
  |||||
511 .....GGAGTGAAGCAGGGGAGTCTTAACTTCGTCGACTCGGAGAA 552

158 HisGlyThrValPheSerSerGluLysLysThrIleLeuMetAlaG1 174
  |||||

```

```

553 GGAGAAACCTTCCTCGCTTCAGATATTCGCGCAATATCTCTTTACACGAA 602
174 uGluMet..... 176
603 AAAGATTATTGTTCTTGATGACGGGGAATAGCGGACCTGACTCCCGACA 652
177 .....AsnLeuAsp..... 179
653 CTGTGAACATTTTACAACCTTTGAGGAGAGCCCGCTTTCAAAGGAAGTAATG 702
180 .....LeuGlyLeuAspLysArgThrIleGluHisTy 190
703 ATTACGCCCTGGGATCTGTCTCGGAAAAGGGTGGTTTAAACACTT 752
190 rValAspLeuGlnTyrValProGluPro.....AspThrLeuH 203
753 CATGCTAAAGAGATATACGACGACCCCAAGCCATAAACGACACACTCA 802
203 isAlaGlnIleSerArgLeuGluSer..... 211
803 AGGTTTCTCTCAACCGAAGACGCAATACCTTTAAGTTAAAGACTTC 852
212 .....GlyCysThrAlaThrValArgProGly..... 220
853 AGAAGGTTTAAATAATAGCTCGGACCTCTTACCACGCGGCTTCGT 902
221 ..GlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnL 237
903 CGAAAGTACTGGATAGAGATTT.....GCAGGTGTTCCACACAGAGG 946
237 ysVal.....ValLysGlyLysGluGln 244
947 TAATTTACGCTTCGGAATTCAGGTATCGGACGCTTCCCGTTTCGGACAAG 996
245 AspLeuPheAspArgIleAlaGln..... 252
997 GATATCGTTATCGGAATTTCCCAAGTCAGGAGACCGCTGCACACAAAGTT 1046
253 ..ValLeuGluAspSerValGluLysHisMetArgAlaAspValThrValG 269
1047 TGCCTTCAGTCGCAAGGAAAG.....GGAGCCTTTACCGTGG 1087
269 ly.....SerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaPro 283
1088 GACTCGTAAACGCTAGTGGGAAGTCCATAGACAGGAGTGGACTTTTCC 1137
284 LeuAlaLysArgHisAsnProAspLeu.....LeuThrPh 295
1138 CTTCACACACATCGGGACCCGAAATAGCGTGGCGGCTACAAAGACCTT 1187
295 eThrThrGlyPhe.....G 300
1188 CACCGCACAGTTCACCGCACTCTACGCCCTTTTCGTTAAGGGAAGTACG 1237
300 luArgGluGly.....TyrSerGluValaspValAlaIleGlu 312
1238 AGAGGAAAAATCTAATAAGACTCTTGTAAAGGTTCCATCCTCGTGTAA 1287
313 SerAlaAlaAlaIleGlyAlaGluHisIleValLysIleValSerProG1 329
1288 CAAACACTGAACACC...GCAGAAAGAGTGGAGAGGTACGGGAANGTA 1334
329 uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProv 346
1335 CATGAAAGAAAAACATGCTTTTACCTCGGAAGGTACTTAAATTACCCCA 1384
346 alala.....AspProSerLeuValProLeuTyrPheValAlaAla 359
1385 TAGCGTGGAGGAGCTCTTAAACTTAAAGAAATTTCTTACATACACGCG 1434
360 GluAla.....ArgLysHisValLysValValLeuSerG1 371
1435 GAAGGTTATCCCGCAGGGAGATGAACACGCTCCATAGCCCTCATAGA 1484

```

alignment\_scores:

```
190 rValaspLeuGlnTyrValProGluPro.....AspThrLeuH 203
    : : : : : : : : : : : : : : : : : : : : : :
753 CATGCTAAAGAGATATACGAACGACCCAAAGCCATAAAGCACACTCA 802
    : : : : : : : : : : : : : : : : : : : : : :
203 isAlaGlnIleSerArgLeuGluSer.....                211
    : : : : : : : : : : : : : : : : : : : : : :
803 AGGGTTTCTCTCAACCGAAGACGCAATACCCCTTTAAGTTAAAGACTTC 852
    : : : : : : : : : : : : : : : : : : : : : :
212 .....GlyCysThrAlaThrValArgProGly..... 220
    : : : : : : : : : : : : : : : : : : : : : :
853 AGAAGGGTTTTAATAATACGTCGGGAGCTCTTACCACGCGGGCTTCGT 902
    : : : : : : : : : : : : : : : : : : : : : :
221 .GlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnL 237
    : : : : : : : : : : : : : : : : : : : : : :
903 CGCAAGTAGTACTGGATAGAGAGATT.....GCAGGTGTTCCACAGAGG 946
    : : : : : : : : : : : : : : : : : : : : : :
237 ysVal.....VallysGlyLysGluGln 244
    : : : : : : : : : : : : : : : : : : : : : :
947 TAATTTACGCTTCGGAAATTCAGGTATCGGACGCTTCCCGTTTCGGACAAG 996
    : : : : : : : : : : : : : : : : : : : : : :
245 AspLeuPheAspArgIleAlaGln.....                252
    : : : : : : : : : : : : : : : : : : : : : :
997 GATATCGTTATCGGAATTTCCCAAGTCAGAGACCGCTGCACACAAAGTT 1046
    : : : : : : : : : : : : : : : : : : : : : :
253 .ValLeuGluAspSerValGlnLysHisMetArgAlaaspValThrValG 269
    : : : : : : : : : : : : : : : : : : : : : :
1047 TGCCCTTCAGTCGCAAGGAAAG.....GGAGCCTTTACCGTGG 1087
    : : : : : : : : : : : : : : : : : : : : : :
269 ly.....SerPheLeuSerGlyGlnLeuAspSerThrAlaIleAlaPro 283
    : : : : : : : : : : : : : : : : : : : : : :
1088 GACTCTAAACGTAGTGGGAAGTGCATAGACAGGAGTGGGACTTTTCC 1137
    : : : : : : : : : : : : : : : : : : : : : :
284 LeuAlaLysArgHisAsnProaspLeu.....LeuThrPh 295
    : : : : : : : : : : : : : : : : : : : : : :
1138 CTTACACACATCGCGGACCGCAAGATAGCGTGGCGCTACAAAGACCTT 1187
    : : : : : : : : : : : : : : : : : : : : : :
295 eThrThrGlyPhe.....G.....G.....G.....G.....G 300
    : : : : : : : : : : : : : : : : : : : : : :
1188 CACCGCACAGTTTCCCGCACCTCTACGCCCTTTTCGGTAAGGGAAGTGAG 1237
    : : : : : : : : : : : : : : : : : : : : : :
300 luArgGluGly.....TyrSerGluValAspValAlaAlaGlu 312
    : : : : : : : : : : : : : : : : : : : : : :
1238 AGAGGGAATCTAATAAGACTCTTGAAGAGTTCCATCATCTCGTTGAA 1287
    : : : : : : : : : : : : : : : : : : : : : :
313 SerAlaAlaAlaIleGlyAlaGluHisIleValLysIleValSerProGl 329
    : : : : : : : : : : : : : : : : : : : : : :
1288 CAAACACTGAACACC...GCAGAAGAAGTGGAGAGGTAGCGGAAAGTA 1334
    : : : : : : : : : : : : : : : : : : : : : :
329 uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProv 346
    : : : : : : : : : : : : : : : : : : : : : :
1335 CATGAAAAAAGAAACATCTTTACCTCGGAAGGTACTTAAATTTACCCCA 1384
    : : : : : : : : : : : : : : : : : : : : : :
346 aAla.....AspProSerLeuValProLeuTyrPheValAlaAla 359
    : : : : : : : : : : : : : : : : : : : : : :
1385 TAGCGCTGAGGGAGCTCTTAACCTTAAGAATAATTTCTTACATACACGG 1434
    : : : : : : : : : : : : : : : : : : : : : :
360 GluAla.....ArgLysHisValLysValValValSerGl 371
    : : : : : : : : : : : : : : : : : : : : : :
1435 GAAGGTTATCCCGCAGGAGATGAAGCAGCGTCCCATAGCCCTCATAGA 1484
    : : : : : : : : : : : : : : : : : : : : : :
371 yGluGlyAlaAspGluPheGlyTyrThrIleTyrLysGluProL 388
    : : : : : : : : : : : : : : : : : : : : : :
1485 CGAAAAACATGCCGGTTGTGTA.....                1506
    : : : : : : : : : : : : : : : : : : : : : :
388 euSerLeuAlaPro.....PheGluLysIleProSerProLeu 400
    : : : : : : : : : : : : : : : : : : : : : :
1507 .....ATCGCACCGAAGACAGGGTTTAGAGAGATATCTCTCAACCGTA 1551
    : : : : : : : : : : : : : : : : : : : : : :
401 ArgLysGlyGluGlyLysLeuSerLysValLeuProAspGlyMetLysGl 417
    : : : : : : : : : : : : : : : : : : : : : :
1552 GAAGAGGTTCTCGCAAGAAAGGAGGGTTATTCTGTAGGCTTTAAAGG 1601
    : : : : : : : : : : : : : : : : : : : : : :
417 lysSerLeuLeuGluArgGlySerMetThrMetGluGlu 430
```

```
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-328-111-544

seq_documentation_block:
; Sequence 544, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328.111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 544
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(570)
; OTHER INFORMATION: n = A,T,C or G
; US-09-328-111-544

alignment_scores:
    Quality: 152.00      Length: 198
    Ratio: 1.434        Gaps: 7
    Percent Similarity: 53.535    Percent Identity: 26.263

alignment_block:
US-09-786-474-2 x US-09-328-111-544/rev ..

Align seg 1/1 to reverse of: US-09-328-111-544 from: 1 to: 570

105 GlyGluProIleValValGlyPheHisHisTrpGlyGluSerValValGl 121
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 GTGAGATAATCCTT.....CATCTTATGCCAAGAGGAAATGA 505

121 uHis.....LeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 GCAANCAATTTGTATGTTGGATGCTGTNGCATTTGTTTACTGGATA 455

134 hrlsGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150
    || ||||| ||||| ||||| ||||| ||||| ||||| |||||
454 CTGCCAATAAGAAAGTGTCTCGGGTAGAGATACATATGGAGTCAGACCT 405

151 LeuPheTyrAlaThrThrGluHisGly...ThrValPheSerSerGluLy 166
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 TTCTTTAAGCAATGACAGAGATGGATTTTGGCTGTATGTCAGACGC 355

166 sLysThrIleLeuGluMetAlaGluMetAsnLeuAspLeuGlyLeuA 183
    ||| ||||| : : : : : ||||| ||||| |||||
354 TAAAGTCTTGTACATTGAAGCACTCCCGCACTCCCTTTTAAAGTGG 305

183 sLysArgThrIleGluHisTyrValAspLeuGlnTyrValPro..... 197
    : : : : : ||||| ||||| ||||| |||||
```

```
304 AGCCTTTTCCTCGGACACATGAAAGTTTGGATTTAAAGCCAAATGGC 255
198 .....GluProAspTh 201
254 AAGTTGCATCCGCGGAATGTTAAATATATCATCTCGGGATGAACC 205
201 rLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVala 218
||||| .....
204 CCTGCAGCCCTCTATGACAATGTGGAGAAA.....CTCT 170
218 rgProGlyGlyLeuGluGlnLysArgTyrPheLysProGlnPhePro 234
||||| .....
169 TTCAGGTTTGGATAGAA..... 150
235 ValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIleAl 251
||||| .....
149 .....ACTGTGAAGAACCTCAGGATCCTTTTAAT 117
251 aglnValLeuGluAspSerValGluLysHisMetArgAlaAspValThrV 268
||||| .....
116 .....AATCGTAAAGAACGTTTGATGACAGACAGAAGGA 80
268 alGlySerPheLeuSerClyGlyIleAspSerThrAlaIleAla 282
||||| .....
79 TTGGCTGCCTTTTATCAGGGGCTTGGACTCCAGCTGGTTGCT 36
```

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_comb.seq:US-09-328-111-395

seq\_documentation\_block:

```
; Sequence 395, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 395
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(629)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-395
```

alignment\_scores:  
Quality: 132.00 Length: 170  
Ratio: 1.333 Gaps: 10  
Percent Similarity: 58.235 Percent Identity: 26.471

alignment\_block:

US-09-786-474-2 x US-09-328-111-395 ..

Align seg 1/1 to: US-09-328-111-395 from: 1 to: 629

```
1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
||||| .....
128 ATGTGTGCATTTGGCGCTGTTTGGCAGTGATGATTCCTTCTGCT... 175
17 eValProAlaLeuGluArgAlaLeuProCysMetArg.....HisArgG 32
||||| .....
176 .....CAGTGTCTGAGTGCTATGAAGATTCACACACAG 209
32 LyProAspAlaGlyThrTrpHisAsp.....AlaAspAla 44
||||| .....
210 GTCCA...GATGCATTCGCTTTTGAGATGTCATATGGATACACCACTGC 256
45 AlaPheGlyPheAsnArgLeuSerIleIleAspIleAlaHisSerHisGl 61
||||| .....
257 TGCITTTGGATTTCCACCGCTTGGCGGTAGTTGACCCGCTGTTTGGAA 306
61 nProLeuArg.....CTCTGTACAATGGTGAATCTACAACCTAAGAAG 376
||||| .....
307 GCCAATTCGAGTGAAGAAATATCCGTATTTGTGG..... 340
71 roAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrValGlu 87
||||| .....
341 .....CTCTGTACAATGGTGAATCTACAACCTAAGAAG 376
88 LeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGlyAs 104
||||| .....
377 ATGCAACAGCATTTTGAA.....TTTGAATACCAGACCAAAAGTGA 417
104 pGlyGluProIleValValGlyPheHisHisTrpGlyGluSerValValG 121
||||| .....
418 TGGTGAG...ATAATCCTTCATCTTTATGACCAANGAGGAATTGAGCCA 464
121 luHisLeuArgGlyMetPheGly.....IleAlaIleTrpAspThr.... 134
||||| .....
465 ACCATTTCGNATGGTTGGATGGGTGGTGTGCAATTNGGTTTACTGGGAA 514
135 .....LysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyI1 148
||||| .....
515 ACTGGCCATTANGAAAGGGNTCCTGGGTAAAGAAGATCCCTATGGGGCC 564
148 eLysProLeu 151
|||||
565 NNAACCTTTG 574
```

